Genome Sequencing and Phylogenetic Analyses as a Basis for Molecular Subtyping of Male-Specific (FRNA) Coliphages

by
Stephanie Dianne Friedman

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environmental Sciences and Engineering, School of Public Health

Chapel Hill
2008

Approved by:

Mark D. Sobsey, PhD

Jan Vinjé, PhD

Fred J. Genthner, PhD

Lola V. Stamm, PhD

Frederick K. Pfaender, PhD

Louise M. Ball, PhD

Rachel T. Noble, PhD

Abstract

Stephanie Dianne Friedman

Genome Sequencing and Phylogenetic Analyses as a Basis for Molecular Subtyping of Male-Specific (FRNA) Coliphages

(Under the direction of Mark D. Sobsey)

Monitoring programs for recreational waters utilize indicator bacteria concentrations as predictors of sewage-exposure related illness risks. However, most illnesses contracted through exposure to recreational waters may be of viral etiology. Identifying the fecal sources (non-human vs human) is also valuable information for risk management and source mitigation. Male-specific (FRNA) coliphages are proposed as sensitive enteric viral indicators for source-tracking fecal pollution in environmental waters. Classified as family *Leviviridae* of two genera, *Levivirus* and *Allolevivirus,* and four genogroups (I, II, III, IV) the genogroups provide information regarding animal or human fecal sources. In order to design an assay for molecular identification of specific genogroups, a genomic sequence database of sufficient size must be generated from several FRNA coliphages collected from diverse sources or locations. The complete genome of 21 FRNA strains was sequenced and compared with 11 strains available in GenBank. Sequences of 30 out of 32 FRNA coliphages demonstrated very similar conserved regions, Open Reading Frame positions, amino acid compositions and gene maps when compared to the FRNA reference strains. The sequence of two strains could be placed in a new subcluster of genogroup I and further analysis suggests that these viruses are natural recombinants. Among viruses within each

genogroup, nucleotide sequence similarities ranged from 75-99%, 83-93%, 69-95% and 74-95% for genogroups I, II, III and IV, respectively. Genogroup II lysis protein tree formed a unique branch that was not observed in the full-length nucleotide tree. Thus, both full-length nucleotide and individual protein sequences need to be evaluated when genotyping or phylogenetically clustering these FRNA coliphages. From conserved regions within each genogroup, four genogroup-specific primer sets were designed for a reverse-transcription polymerase chain reaction (RT-PCR) assay. The assay was then evaluated successfully on a panel of environmental FRNA strains demonstrating their usefulness to assess the sanitary quality of recreational waters and provide data identifying and subsequently eliminating the contamination source.

.

Acknowledgments

It has been my privilege to be a member of the UNC Environmental Virology laboratory and to study under the direction of Drs. Mark Sobsey, Jan Vinjé and Fred Genthner. A generous thank you is extended to each of you for your guidance and assistance. A special gratitude to Dr. Genthner for going above and beyond to secure funding for the project and for dealing with the never ending layers of administration. I would also like to thank Drs. Ball, Stamm, Pfaender and Noble for serving on my committee.

The isolation and preliminary characterization of the majority of these phage was performed by Greg Lovelace and David Love (University of North Carolina at Chapel Hill). A portion of this work was funded by a RARE grant secured through the efforts of J. Paar, III, EPA Region 1 Laboratory, North Chelmsford, MA.

Table of Contents

Table of Contents vii

List of Tables ix

List of Figures x

I. Introduction 1

II. Background 5

III. Research Objectives 9

IV. Literature Review 10

References 49

V. Gene Mapping and Phylogenetic Analysis of the Complete Genome of 30 ssRNA Male-Specific Coliphages of the Family *Leviviridae* 56

Introduction 57

Purpose 59

Materials and Methods 60

Results 69

Discussion 88

Summary 94

References 95

VI. Genomic sequences of two novel *Levivirus* FRNA coliphages (*Leviviridae):* Evidence for recombination? 99

Introduction 100

Purpose 103

Materials and Methods 104

Results 112

Discussion 124

Summary 129

References 130

VII. A Reverse Transcription-PCR Assay to Distinguish the Four Genogroups of Male-Specific (F+) RNA Coliphages 133

Introduction 134

Purpose 136

Materials and methods 137

Results 148

Summary 153

References 154

VIII. Overall Discussion 157

IX. Summary and Conclusions 171

X. Recommendations for Future Research 173

References 179

Appendix A1 182

Group I 182

Group II 201

Group III 213

Group III QB-like 235
Group IV 253
Appendix B1 271
Group I 271
Group II 279
Group III 285
Group IV 294
Appendix C 303
Group I and JS Nucleotides 303
JS Amino Acids 325
Group I and JS Amino Acids 329
Group I and JS Replicase Nucleotides 341

List of Tables

Table **Page**

Table 4.1 Bacteriophage classification. 14

Table 5.1 Open Reading Frames positions for *Leviviridae.* 70

Table 5.2 Leviviridae nucleotide percent similarity 72

Table 5.3 Start condons and Shine-Dalgarno sequences 76

Table 5.4 *Leviviridae* number of amino acids per gene 79

Table 5.5 Comparison of genomic traits. *Allolevivirus* and *Levivirus* 85

Table 6.1 JS strains and genogroup I 115

Table 6.2 Pairwise comparison JS strains and genogroup I 116

Table 6.3 Pairwise comparison of amino acids for JS and genogroup I. 118

Table 6.4 Pairwise comparison of replicase protein JS and genogroup I. 119

Table 7.1 Sources of *Leviviridae* strains. 138

Table 7.2 *Leviviridae* accession numbers 142

Table 7.3 Genogroup-specific primer sets. 149

List of Figures

Figure **Page**

Figure 4.1 Somatic and male-specific coliphage hosts. 26

Figure 4.2. Schematic of male-specific coliphage classification. 29

Figure 5.1. Flow-chart of sequencing methods. 62

Figure 5.2 Phylogenetic analysis of *Leviviridae* nucleotide sequences. 73

Figure 5.3 SimPlot nucleotide similarity and genome organization. 73

Figure 5.4 (A) Phylogenetic analysis of each protein: Levivirus I, II. 86

Figure 5.4 (B) Phylogenetic analysis of each protein: Allolevivirus III, IV. 87

Figure 6.1 SimPlot analysis of genome nucleotide profile strain MS2 and DL52. 120

Figure 6.2 SimPlot analysis of genome nucleotide profile strain MS2 and fr. 120

Figure 6.3 SimPlot analysis of replicase amino acid profile strain MS2 and DL52..... 121

Figure 6.4 Phylogenetic analysis of replicase amino acids groups I, II and JS. 122

Figure 6.5 JS strains replicase frame-shift. 123

Figure 7.1 Gel electrophoresis of FRNA phages using one-step RT-PCR. 143

Figure7.2 Alignment and primer design of FRNA phages for one-step RT-PCR. 144

I. Introduction

A link between waterborne transmission of disease and sewage was first observed in 1854 by the historical achievements of Dr. John Snow. The observations by Dr. Snow pioneered the sciences of epidemiology, preventative medicine and public health intervention. His observations of the cholera outbreak led to the removal of the Broad Street pump handle in London followed by a rapid, subsequent decline in cholera-related deaths (Snow, 1855).

Sewage disposal into marine and freshwater systems has occurred since the establishment of community populations. In the USA and most developed countries treated wastewaters are commonly discharged into aquatic environments, including both drinking-water resources and recreational waters. Fecal contamination introduced into waterways may potentially contain both human and/or zoonotic pathogens, including various enteric and respiratory viruses. A virus- infected individual may fecally shed viral particles for weeks at levels up to 10^5 to 10^{12} per gram of stool (Bosch, 1998; Griffin et al, 2003; Gerba, 2000), depending upon the viral species or type. The majority of the shed viruses are thought to enter the aquatic environment primarily from sewage discharge, but additional routes include solid waste applications (Bosch. 1998), and shedding during recreational bathing (Gerba, 2000). Although treating wastewater by conventional primary and secondary treatment and terminal disinfection reduces the majority of fecal pathogens, resistant bacteria, parasite cysts, oocysts and spores and some viruses may not be removed adequately. Besides

municipal wastewater discharges, inputs such as stormwater runoff, septic-tank seepage from on-site systems, agricultural runoff, urban runoff and other fecal waste sources can enter aquatic environments. Other naturally occurring, non-pathogenic and potentially pathogenic viruses are already present in aquatic environments, thus adding to the complexity of the viral ecology of aquatic systems.

To minimize adverse impacts and to protect the public health and aquatic environments, management systems based on fecal indicators of microbial origin were implemented as a "warning flag" and a means by which to estimate fecal contamination based on direct microbial measurements. The World Health Organization (WHO; Ashbolt et al, 2001) defines microbial indicators of public health concern by one of three groups:

> "(1) process indicator - group of organisms that demonstrates the efficacy of a process, such as total heterotrophic bacteria or total coliforms for chlorine disinfection
> (2) fecal indicator - a group of organisms that indicates the presence of a fecal contamination, such as thermotolerant coliforms or *E. coli*. Hence, they only infer that pathogens may be present (3) index and model organisms - a group or species indicative of pathogen presence and behavior, respectively. For example, *E coli* as an index for Salmonella; F-RNA coliphages as models of human-enteric viruses."

Criteria for an ideal fecal indicator selection are as follows: (1) consistently present in feces at higher concentration than those of pathogens; (2) cannot/should not replicate outside the intestinal tract; (3) easily detected and quantified; (4) non-pathogenic; (5) at least as resistant as pathogens to disinfection treatments and environmental conditions/inactivation rates; (6) indicator concentration in water is quantitatively associated with potential risks to human populations, typically from enteric illness; and (7) applicable to all water types (marine, freshwater, estuarine).

Existing bacteriological culture methods for determining fecal contamination in water

provide quantitative estimates of *E. coli* and enterococci. These methods are neither real-time nor do they provide information regarding source. The best that current methods can do is indicate that possible fecal contamination occurred within the last 24 hours. To minimize risks to human health, resource managers and human health advisors need an early-warning indicator that will enable them to assess the sanitary condition of waters in real-time or at least shortly after the sample is collected for analysis. An additional limitation of the current indicators, enterococci and *E. coli*, is that they do not correlate with the presence and concentrations of all potential water-borne pathogens (Griffin et al., 2003). Most importantly, current EPA recreational water-quality criteria using two bacterial indicators have little or no correlation to the presence and concentration of human pathogenic viruses. To date, no viral indicator has been mandated for regulatory purposes in recreational waters in the USA.

Male-specific coliphages have been suggested as a viral indicator for: (1) fecal contamination (Osawa, 1981; Furuse, 1983), (2) enteric bacterial contamination (Gerba, 1987), (3) enteric viral contamination (Grabow, 2001; Leclerc et al, 2000) and (4) risks of gastro-intestinal illness from recreational water exposures (Colford et al., 2007). Male-specific coliphages, specifically the ssRNA *Leviviridae* family, are superficially indistinguishable from most human enteric viruses (Grabow, 2001), occur in higher numbers in sewage and wastewater effluents than viral enteric pathogens (Grabow, 2001), their presence implies the presence of pathogenic viruses (Grabow, 2001), and, in a majority of cases, they provide human/animal fecal-source specificity (Vinjé et al, 2004; Cole et al, 2003; Furuse, 1987; Schaper et al, 2002; Scott et al, 2002; Stewart, 2002; Long et al, 2005).

The focus of this research was to (1) generate a nucleotide sequence database by

sequencing at least three to five strains from each FRNA coliphage genogroup (I, II, III and IV), (2) develop and analyze for identification of preferred targeted genomic regions the sequence database representing environmental and prototype strains and (3) based on the primers identified in step 2, design and validate a molecular assay to detect and ultimately subtype the different genogroups. To develop a genetic database, 19 FRNA strains were sequenced. In addition, two new undescribed *Levivirus* strains were sequenced. In addition, a one-step reverse transcription polymerase chain reaction (RT-PCR) detection method for the four FRNA coliphage groups (I, II, III and IV) was designed to distinguish human vs. animal fecal source. Highly precise forward and reverse genogroup-specific primers were designed based on a total of 30 FRNA sequences containing several strains from all four genogroups.

II. Background

To develop valid fecal-indicator criteria based on credible epidemiology design, US EPA undertook a series of marine and freshwater public-beaches studies (EPA-600/1-84-004; EPA-600/1-80-031). Objectives were to assess the mathematical relationship between microbiological indicator concentrations in bathing water and illness rates resulting from recreational water exposure (swimming), in order to correct for perceived deficiencies in US Public Health Service studies conducted before 1972. Additional goals were to provide a statistically sound study outcome correlating the concentration of the best bacterial indicator with magnitude of health effects in bathers, resulting in specific bacterial concentration-health risk outcome relationships associated with swimming in sewage-contaminated waters. Two key US EPA documents were published from these epidemiological studies "Health Effects Criteria for Fresh Recreational Waters" (EPA-600/1-84-004) and "Health Effects Criteria for Marine Recreational Waters" (EPA-600/1-80-031). Epidemiological data supported the use of *E. coli* and enterococci as primary fecal indicators because they were associated with statistically significant increased gastrointestinal illness rates to swimmers/bathers at increasing concentration in water. Subsequently, EPA guidelines for recreational waters were established in the 1986 "Ambient Water Quality Criteria for Bacteria in Recreational Waters."

To overcome deficiencies in the 1986 US EPA guidelines, the agency was compelled to implement the Beaches Environmental Assessment and Coastal Health Program (BEACH)

Act of 2000. The BEACH Act was an amended Section 303 of the Clean Water Act. The focus was to improve public health and recreational water quality programs by 1) strengthening beach testing and standards, 2) providing faster testing methods, 3) predicting fecal pollution, 4) better defining the criteria as to which fecal indicators and water quality standards are based, 5) investing in health and methods research and 6) informing the public (EPA, 2003).

There are limitations associated with the two current bacterial fecal indicators for recreational water quality, enterococci and *E. coli*. They do not provide a complete assessment of the sanitary quality of water. Their presence and densities do not correlate with all potential water-borne pathogens causing adverse environmental effects (Griffin et al., 2003). For example, most illnesses contracted by swimmers appear to be of viral etiology. Environmental and health risks associated with microbes can be as much a consequence of viral contamination as it can of contamination by any other microorganisms. Both microbial and epidemiological evidence suggests that the current EPA recreational water-quality criteria using two bacterial indicators have little or no correlation to the presence of pathogenic viruses or to health-risks from non-point source fecal contamination. In microbial water quality studies conducted in Florida, indicator bacteria counts satisfied EPA criteria limits in the presence of detectable human pathogenic viruses (Griffin et al., 2003; Griffin et al., 1999). In other studies, waters satisfying *E. coli* standards for shellfish harvesting often harbored shellfish that contained human enteric viruses (Dore et al., 2000). In laboratory studies, bacterial indicators have been erroneous predictors of virus survival because viruses survived longer and were inactivated at slower rates than were bacteria (Burkhardt III et al.,

2000; Sinton et al., 1999; Sinton et al., 2002). Even when current bacteriological standards are met in recreational waters, risks to human health may be posed by viruses.

In other studies, *E. coli* and enterococci were detected in environmental niches of tropical or temperate climates (Hawaii, Guam and Florida) where there was no evidence of human fecal contamination. Such environmental reservoirs of these bacteria led to high indicator counts that were not associated with fecal contamination (Hardina and Fujioka, 1991; Roll and Fujioka, 1997; Byappanahalli and Fujioka, 1998; Fujioka and Byappanahalli, 2000; Solo-Gabriele HM et al., 2000; Genthner et al., 2005). In some cases, elevated bacterial indicator counts exceeding EPA water-quality criteria were influenced by soil run-off and not a result of sewage input (Byappanahalli and Fujioka, 2004).

The apparent weaknesses in current bacterial indicators and the need to develop better, faster methods to accurately measure fecal contamination and determine source (animal vs human) supports the proposed use of a suite of indicator species, including at least one viral indicator for more specific and timely water-quality public health assessments. An attractive viral candidate is male-specific (F+), ssRNA bacteriophages (FRNA coliphages). They are viruses that are morphologically similar to most enteric viruses (Grabow, 2001), their presence in sewage and wastewater-impacted waters implies the presence of pathogenic viruses (Grabow, 2001), they are associated with "fresh" fecal contamination, they are easily detected in environmental samples, they survive disinfection treatments similar to other pathogenic viruses (Duran et al, 2003, Nappier et al, 2006) and they occur in higher numbers in sewage and wastewater effluents than viral enteric pathogens (Grabow, 2001). FRNA phages have been isolated from treated wastewater samples in the absence of fecal coliforms (Stewart, 2002), they provide insight as to contamination source (Vinjé et al., 2004; Cole et

al., 2003; Furuse, 1987; Schaper et al., 2002; Scott et al., 2002; Stewart, 2002; Long et al., 2005) and the potential exists to develop and use rapid nucleic-acid based molecular detection for real-time public-health risk management.

To design and develop a robust molecular genogroup-specific assay, the analytical system must be based on a genetically-representative sequence database. To identify appropriate genomic regions to target for primer design and amplification the use of only the few FRNA phages that have been fully sequenced over the past few decades is inadequate. As of 2007, eleven full-length genome sequences were available in GenBank (GenBank/EMBL/DDBJ). The use of more comprehensive and representative genetic sequence data for these coliphages would provide greater assurance of the development of a reliable molecular detection system based on RT-PCR amplification.

III. Research Objectives

Specific Goal:

1. To develop and validate a rapid genetic typing assay for the detection and characterization of FRNA coliphages from various geographical locations and sources representing each genogroup (I, II, III, IV).

Study Design:

1. Increase the full-genomic FRNA sequence database by selecting and sequencing representative strains from each genogroup.

2. Conduct phylogenetic analyses, identify nucleotide similarities, locate Open Reading Frame (ORF) positions and gene locations from these sequences.

3. Perform bioinformatics analysis to reveal protein domains, family conservations and discrete amino acid sequence features, or motifs, by Pfam and PROSITE patterns.

4. Identify optimal target regions for the development of a one-step reverse transcription-PCR assay specific for each genogroup.

5. Design and validate the one-step RT-PCR assay specific for each genogroup.

6. Identify optimal target regions to design primers and probes for the development of a real-time PCR assay specific for each genogroup.

7. Design the real-time assays to detect and discriminate the four *Leviviridae* genogroups, I, II, III and IV.

IV. Literature Review

Coliphage History

Between 1915-1917, bacterial viruses or bacteriophages, were first noted independently by Frederick William Twort and Felix Hubert d'Hérelle. Approximately 20 years earlier, plant viruses (1892 by Ivanowski) and animal viruses (1902 by Loeffler and Froesch) had been described (Duckworth, 1987). During the late 1930's, Emory L. Ellis isolated phages from Pasadena, CA, sewage by using *E. coli* as a host bacterium. In 1942, bacteriophages were first observed using electron microscopy (Ackermann, 2006) by Tom Anderson (www.asm.org). Bacteriophage lambda was discovered in 1951 by Esther Lederberg while observing a lysogenic phage isolate from *E. coli* (Ackermann, 2006). Throughout the 20th century, the culmination of phage research laid the foundation of modern molecular biology.

An RNA phage, strain f2, was isolated and characterized in 1961 by Loeb and Zinder (Loeb and Zinder, 1961; Zinder, 1965; Furuse, 1987; van Duin, 1988) and in 1959, Dr. A.J. Clark isolated the most well-studied RNA phage, strain MS2 (personal communication). These ssRNA coliphages infect gram-negative bacteria expressing a sex-pili (F+), are sensitive to RNase and are collectively known as male-specific phages (FRNA). Using electron microscopy, Bradley (1964) first described RNA-containing phage morphology as an icosahedral form. In 1976, strain MS2 was the first virus to be completely sequenced

(Fiers et al., 1976). Genomic sequences could now be compared to serological and physicochemical typing.

Taxonomy

Early taxonomic classification of phages, between 1920s and 1930s, was based on bacterial host specificity. In the 1940s and 1950s, the advent of electron microscopy propelled phage taxonomy to morphological descriptions (Nelson, 2004). Techniques to isolate nucleic acids and determine the genome composition, i.e., dsDNA, dsRNA, ssRNA, ssDNA, provided a system of viral taxonomy based on nucleic acid type and morphology published by Lwoff, Horne and Tournier in 1962 (Ackermann, 2006). Bradley proposed a classification scheme based on basic phage features, i.e., nucleic acid type (DNA or RNA, ss or ds, circular or linear), capsid morphology, tailed or filamentous phages, enveloped or non-enveloped, etc. By the late 1960s to early 1970s, the International Committee on Taxonomy of Viruses (ICTV) formalized phage classification into six genera based on Bradley's proposed scheme. As new phage types are discovered, they are added to the ICTV classification of one order and 13 families. The largest viral group is the bacteriophages, with the predominant phage type having dsDNA and a smaller number of phage types having ssRNA (Ackermann, 2006).

Genomic or partial genomic sequencing has become the norm in the modern virology laboratory displacing the historical electron microscopy visualization of phage structures. Rohwer and Edwards proposed a sequence-based taxonomic system based on "signature genes" elucidated in phage genomes comparable to the 16S rDNA classification in bacteria (Rohwer and Edwards, 2002). Since no single protein marker or motif was conserved throughout the range of phage genomes, phage relatedness was based on common

characteristics. The more individual characteristics that were shared such as protein sequences, nucleic acid form, etc, the stronger the relationship. Protein-distance programs calculated the number of amino acid changes from protein-to-protein and were a resourceful tool in determining the fine parameters of the phage proteomic tree. The resulting proteomic tree was compatible with the ICTV arrangement (Rohwer and Edwards, 2002).

Male-specific phages belong to either the *Leviviridae*, single-strand RNA, or *Inoviridae*, a single-stranded DNA family (Sobsey et al., 2005) (Fig 4.1, Fig 4.2). Male-specific RNA phages (FRNA) are non-enveloped, positive sense, single-strand RNA (ssRNA) contained within a 26 nm diameter icosahedral-shaped capsid (Buchen-Osmond, 2003). The *Leviviridae* family (Table 4.1) comprises two genera, *Levivirus* and *Allolevivirus*. These genera are further divided into four genogroups (I, II, III and IV). *Levivirus* are genetically divided into genogroups I and II, and *Allolevivirus* are subdivided into genogroups III and IV.

Leviviridae FRNA phages were initially grouped primarily according to their serological properties in that anti-phage sera prepared against specific phage strains could neutralize some of the other closely-related phages (Watanabe et al., 1967; Miyake et al., 1971). Other methods of grouping were through membrane filtration, elution patterns and buoyant densities (Miyake et al., 1969). Until 1969, three phage groups existed, groups I, II and III and subgroups a,b and c in group III. With the isolation and characterization of strains SP and FI, serological assays indicated that these two strains did not cross-react to anti-sera from groups I, II or III. A new group IV and group V were proposed for strains SP and FI, respectively (Miyake et al., 1969; Sakurai et al., 1988). Eventually, four major groups, I, II, III and IV and subgroups a,b, c and d in group III and subgroups a and b in

group IV were assigned to FRNA phages based on template specificity of RNA replicase (Miyake et al., 1971; Miyake et al., 1973).

Based on a limited number of complete sequences four FRNA genes could be identified (reviewed by Bollback & Huelsenbeck, 2001). These genes code for an assembly or maturation protein, capsid protein, lysis protein and replicase protein in the *Leviviruses* whereas in *Allolevivirus*es, the lysis protein is replaced by a read-through protein. Each levivirus virion contains one molecule of positive sense ssRNA, 180 copies of the capsid or coat protein, one copy of the maturation protein and, in alloleviviruses, approximately 15 copies of the read-through protein (Weber & Konigsberg, 1975; van Duin, 2000; van Duin and Tsareva, 2006). Read-through protein synthesis occurs at a rate of about 6% and, although the exact function is unknown, the combination of the read-through protein with the maturation protein is required for an infectious viral stage (van Duin and Tsareva, 2006). The *Leviviridae* maturation protein is needed for viral infection and virus particle maturation (Olsthoorn et al., 1995) whereas the numerous coat proteins in the virion are used for assembly of phage progeny (Klovins et al. 1997; Weber and Konigsberg, 1975). A single, small (< 100 amino acids), hydrophobic peptide, the lysis peptide, is responsible for cell lysis at the end of the infection cycle in the *Levivirus* genogroups (van Duin, 1988). No distinct lysis protein is present in *Allolevivirus*es; lysis is mediated by the maturation protein (Karnik and Billeter, 1983). The replicase in both genera, also known as viral RNA-dependent RNA

polymerase, is required in small amounts early in the infection process. Twenty minutes after infection, synthesis of this protein ceases (van Duin, 1988).

Table 4.1 Bacteriophage classification.

Family	Number of Genera	Morphology	Nucleic Acid	Enveloped
Cystoviridae	1	isometric	ssRNA, L, S	Yes
Leviviridae	2	icosahedral	ssRNA, L	
Corticoviridae	1	isometric	dsDNA, C, T	
Microviridae	4	icosahedral	ssDNA, C	
Tectiviridae	1	icosahedral	dsDNA, L	
Inoviridae	2	rod	ssDNA, C	
Myoviridae	6	contractile tail	dsDNA, L	
Siphoviridae	6	noncontractile tail	dsDNA, L	
Podoviridae	3	short tail	dsDNA, L	
Lipothrixviridae	1	rod	dsDNA, L	Yes
Rudiviridae	1	rod	dsDNA, L	
Plasmaviridae	1	pleomorphic	dsDNA, C, T	Yes
Fuselloviridae	1	lemon-shaped	dsDNA, C, T	Yes

ss - single stranded; L - linear; S - segmented; C - circular; T - superhelical; ds - double stranded

Ecology and Source Specificity

Male-specific ssRNA phages, or FRNA, strains MS2 (isolated by A. Clark), R17 (isolated by Paranchych and Graham) and f2 collected in the late 1950's to early 1960's from the United States were typed into group I. Throughout the 1970's, K. Furuse, I. Watanabe and colleagues conducted a systematic survey by isolating several thousand ssRNA coliphage strains from sewage in domestic drains, feces from humans and various animals, municipal raw sewage, river water, seawater, irrigation, pond or lake waters from across the globe. They classified these phages into four groups, I, II, III and IV, based on antisera testing or sero-typing and by physicochemical characteristics such as a cesium chloride density profile. Some strains were further subdivided into three to seven subgroups. Group III phages were

predominant in the southwest islands of Japan, the Philippines, Indonesia, Taiwan and Singapore and group II phages were most abundant in mainland Japan (Furuse et al., 1978; Miyake et al., 1971). Domestic drains located in Asian countries had very few FRNA from groups I and IV. Phages isolated from domestic drainage in Korea were from groups II and III whereas southeast Asia phages were predominantly group III (Furuse, 1987). FRNA phages isolated from sewage water in Australia were from group II and, in the USA, groups II and III FRNA phages were collected (Furuse et al., 1975).

FRNA phage isolation frequencies from various countries varied greatly. Only 2.5% of phages collected from sewage in a study in Peru were FRNA, 5% were FRNA in Brazil and by comparison, FRNA were 38% in Taiwan and 30% in Japan (Furuse et al., 1975). The FRNA phages from sewage samples in Brazil and West Germany belonged to group I exclusively. It was unknown whether or not these sewage treatment plants received slaughterhouse waste (Furuse, 1987). Furuse states "it can be reasonably assumed that group I phage observed in raw sewage from treatment plants are most likely introduced from animal sources, and group II and group III phage from human sources." A separate study in Brazil found that thirteen out of 353 sewage and/or fecal samples contained FRNA phages, or approximately 4% and of these thirteen, five were typed as group I, two as group II and six as group III (Miyake et al., 1973).

Out of 93 phages isolated from the Bangkok, Thailand collection, only eight were FRNA phages. Three independent samples were collected from river water, two samples were collected from sewage and no FRNA phages were isolated from stool samples. Of the eight phages, one strain was group I whereas three and four strains were typed as groups II

and III, respectively (Aoi et al., 1972).

Domestic sewage samples in Korea had 56% RNA phages belonging primarily to groups II and III along with 4 group I phages (Osawa et al., 1981). Taiwan sewage and fecal samples included 20% FRNA with 2 group IV, 22 group III, 8 group II and 6 group I isolates (Miyake et al., 1971).

Furuse and colleagues continued to further explore the distribution patterns of the four genogroups of FRNA phages in samples from the following sources: 1) gastrointestinal contents of cows and pigs, 2) feces of domesticated animals (pigs, horses, cattle and fowl), 3) human feces 4) animal feces and 5) sewage obtained from slaughterhouse treatment plants. FRNA groups II and III were isolated in almost equal proportions from human feces whereas the gastrointestinal tract of swine harbored groups I and II. Group I was isolated from feces and gastrointestinal contents from all other animals. Slaughterhouse samples were predominately group I along with a few isolates from group II. In theses studies, group III was only isolated from human subjects and from no other host organism. Group IV displayed the greatest habitat diversity as they were isolated from feces of animals, humans, domestic drainage and raw sewage (Furuse, 1987; Osawa et al., 1981). The Furuse study concluded group I phages were not found in humans or domestic drainage sewage but only in animals or

slaughterhouse sewage. However, their results conflict with the fact that group I strains MS2 and R17 were isolated from municipal sewage systems located in the USA.

FRNA phages isolated in The Netherlands from fecal samples obtained from humans and domesticated animals were serologically typed. As in the Furuse studies, groups I and IV phages were isolated exclusively from animals whereas human sources harbored groups II and III (Furuse, 1987)

Groups II and III were isolated from human feces in FRNA distribution studies from South Africa and Spain (Schaper et al., 2002). In contrast to earlier reports of groups II and III occurring only in human feces and I and IV in animals, the Schaper study isolated group II from cattle and swine and group II and III from poultry. This study casts doubt on the absolute association of phage genotype and source-specificity, but emphasized the observation that group I has not been isolated from human feces. Nonetheless, group I has been isolated in domestic sewage and whether or not animal waste was present in these municipal treatment facilities is a question that cannot be easily answered.

The lack of consistent numbers of FRNA in human feces and the discrepancies of phage types harbored in various animals does not minimize the fact that FRNA are isolated from sewage in numbers ranging from 10^2 - 10^4 PFU/ml (Leclerc et al., 2000). From multiple samples over time, FRNA were isolated from raw sewage at an average of 4.2×10^4 PFU/L (Brion et al., 2002).

FRNA Coliphages as Indicators of Fecal Pollution

FRNA coliphages have been recommended as a possible indicator of enteric viruses as their presence indicates fecal pollution from either humans or animals. A highly significant statistical correlation was observed between the presence of FRNA and

enterovirus virus concentrations in river water, coagulated effluent, chlorinated and UV-irradiation effluents, coagulated river water and lake water but not in raw sewage or biologically-treated sewage (Havelaar et al., 1993). FRNA was proposed as a viral indicator in recreational waters due to the strong correlation between FRNA and enteroviruses. Limitations to this study were that only freshwater, not marine or other environmental water bodies were studied, and only a small geographical area was evaluated (Havelaar et al., 1993).

Marine, freshwater and estuarine waters were selected to study the relationship between coliphages, their *E. coli* host and a few pathogenic bacteria. Multiple sampling stations were located near domestic and industrial sewage discharges. A correlation between male-specific coliphage concentration and *E. coli* concentration were found to be dependent upon the direction and distance from the effluent plume. A greater statistical relationship was noted for male-specific coliphages and the pathogens *Salmonella, P. areuginosa* and *C. albicans* when compared to fecal and total coliforms (Borrego et al., 1987). In this study, FRNA and FDNA phages were not resolved separately.

Drinking water sources were analyzed for the presence of male-specific phages, somatic phages and *Bacteroides fragilis* phages for 30 months. Bacterial indicator assays, total coliforms and fecal coliforms, were also evaluated to determine if a relationship existed between the bacterial and viral indicators. *B. fragilis* HSP40 host was used to select the human-specific phage. Throughout the survey, total or fecal coliform positive sites were also positive for at least one or all three phage types. In some instances, coliform negative sites were bacteriophage positive. When the data was tabled as frequency of indicator organisms by year, F+ specific phages had the highest percentage of positive samples two out of three years. This study did not differentiate between FRNA and FDNA male-specific phages.

Phage *B. fragilis* was isolated less frequently two out of three years. Male-specific phages were suggested as an indicator similar to total coliforms and phage *B. fragilis* was similar to fecal coliforms as a measure of fecal pollution (Armon and Kott, 1995).

For a period of two years, raw sewage and surface water samples were surveyed once or twice per week from sites reflective of different land use and agricultural areas. Following double-agar overlay and RNase sensitivity, plaques were genotyped by hybridization using Hsu's method (1995). FRNA phages were the most abundant F+ phage collected in surface waters (67% out of 105 samples) and sewage (87.5% out of 288) samples. Surface waters were positive for type I (81%) and only one sample had type III, suggesting these surface waters were not influenced by human-impacted effluent. However, type III was the predominant genotype as 57% were isolated from raw sewage (Brion et al., 2002). During this study, type III FRNA were recovered more frequently than the other genotypes when a group of male campers were staying near the sampling site. Following their departure, the presence of type III declined and was no longer detectable after a week. Source-tracking with type III may indicate sewage contamination occurred within the past 7 days (Brion et al., 2002).

A ratio of FRNA to FDNA densities were compared from samples collected in animal feces, municipal wastewater facilities, in potentially impaired surface waters and in agricultural livestock wastewaters. To establish a correlation between FRNA genogroups and sample source, FRNA isolates were serotyped. Background samples were defined as an upstream or background site that was sampled concurrently with the surface water sample. FRNA phages were isolated from bovine waste (18%), swine wastewater (50%), gull (96%), goose (100%) and municipal/human wastewater samples (23%). In comparison, FDNA

isolation frequencies were bovine waste (82%), swine wastewaters (50%), gull (4%), goose (0) and human wastewater (77%) FDNA. The remaining animals, horse, buffalo, cat, cormorant, rooster, dog, llama, donkey and pig did not yield F+ coliphages. In animals, goose had the highest percentage of group I (approximately 98%) followed by swine (51%) and cattle (30%). Groups II and III were only isolated from human (50% group II, 15% III), swine (5% II, 22% III) and bovine (15% II, 0 III). Both FDNA and FRNA phages were isolated from surface waters impaired by human, swine, bovine and background sources. Group I FRNA were predominant in background surface waters (97%) whereas bovine-impacted surface waters had 82% group I followed by 75% group I from human-impaired waters. Human land use sites had the greatest percentage of group II isolates (12%) with a low group II recovery from background (2%) and bovine-waste sites. A correlation of FRNA source-associated genogroups was confounded by differential survival rates for each genogroup and/or strain. The authors concluded that there was a statistically significant link of group II FRNA associated with human-land use sites (Cole et al., 2003).

The US EPA sponsored a field validation of Methods 1601 and 1602 to detect coliphages, both male-specific and somatic, in groundwater systems. In addition, fecal indicators *E coli*, enterococci, total coliforms, *Clostridium perfringens* spores and the presence of enteric viruses (enterovirus, hepatitis A, norovirus, rotavirus and adenovirus) were monitored monthly for one year. Male-specific coliphages, but not specifically FRNA, were recommended by the proposed Groundwater Rule as an alternate fecal indicator as they were detected more often in groundwater sources than somatic coliphages by Method 1602. The study concluded that the use of both a bacterium indicator and a coliphage increased the predictability of fecal contamination in groundwater wells (American Water Works

Association., 2004).

Distinct geographical locations (New Mexico, Massachusetts, Connecticut, Michigan, the Carolinas and Southern California) across the USA were surveyed in an ecological study of FDNA, FRNA and somatic phages. Phage concentrations were compared to fecal coliform counts. Viruses were isolated and enumerated on double-agar overlay and the appropriate phage-specific *E. coli* host. Further separation of phage types were determined by RNase sensitivity followed by serotyping. RT-PCR methods followed by Reverse Line Blot (RLB) methods were used to genotype the phages. Direct fecal samples did not have detectable levels (<3 PFU/gram feces) of male-specific phages in feces other than chicken litter, gull and goose. Only FRNA groups I and IV were detected in these avian species. Cow lagoon isolates typed to group I, whereas hog lagoon predominantly typed group I (32%), 3% group IV along with 18% group III. Hog lagoon had the highest percent of group III FRNA. Wastewater influent and effluent were predominately group II with 12 and 15%, respectively. In addition, wastewater influent had 6% group III, 2% I and 1% IV of FRNA phages. Interestingly, septic water samples did not contain FRNA, but had 100% FDNA strain M13. Grazing animal feces contained large numbers of somatic phages. It was concluded that FDNA M13-like was most prevalent in wastewaters, FDNA were detectable in fecal samples and the link between FRNA group III as a human-associated effluent was not absolute (Long et al., 2005).

To add to the body of information on FRNA occurrences from different sources, phages were collected from wastestreams at two hospitals, a cattle feedlot, pig farm and a poultry farm. Environmental river water samples were taken adjacent to farming fields. Male-specific phages were selected on *S. typhimurium* WG 49 host by double-agar overlay,

transferred and fixed onto a membrane and hybridized with the Beekwilder et al., (1996) designed probes. Hybridization probes detected groups II and III in hospital wastewaters, groups III and IV from swine waters, groups I, III and IV from poultry and groups I and IV from cattle wastes. As a whole, river water samples contained all four genogroups. However, seven of the individual river samples were positive for only genogroup II. Analysis of FRNA from an assortment of sources led the authors to conclude that group III was not necessarily specific for human excreta, but the trend on the specificity of groups I and IV for animal sources and II and III for human sources supported previous findings (Sundram et al., 2006).

The ambiguous association of swine isolates of FRNA coliphages periodically grouping as type III led to the hypothesis that perhaps a refined genogrouping system could clarify if sub- or unique group III clusters existed. Forward and reverse primers for reverse-transcription polymerase chain reaction (RT-PCR) were designed to the 5' untranslated region spanning into the maturation gene region in group III phages. Primers were based on known group III sequences to strains MX1, M11 and Qβ. Isolates were first classified to group III by hybridization (Hsu et al., 1995; Beekwilder et al., 1996). Thirty-two type III coliphage strains were isolated from swine lagoons, surface and wastewater sources from North Carolina and South Carolina. RT-PCR amplification was performed on those isolates testing positive for group III. RT-PCR products of about 567 nucleotides were sequenced and phylogenetic trees were generated from these sequences. Phages isolated from North Carolina lagoons, surface and wastewaters grouped as Qβ-like whereas those from South Carolina were closely related to prototype strain M11. In this case, group III isolates could not be genetically separated as human vs swine clusters. The authors noted genetic similarity

evidence from human and swine hepatitis E virus populations being consistent with their finding for FRNA phages (Stewart et al., 2006).

Ninety-six surface water stations on the State of South Carolina's impaired list were selected as study sites to quantify FRNA and somatic coliphages. Typing FRNA occurred by serological and/or nucleic acid hybridization methods. Fourteen of these sites identified FRNA genogroups II (5%) and III (1%) while the majority of isolates typed as group I (94%). Direct wastewater samples typed 73% group III, 14% II and 11% group I. Isolates collected and typed from a swine lagoon had 70% group I, 19% group III and 6% group IV. Chicken litters contained approximately equal amounts of group I and IV FRNA. The presence of groups II and III from the 14 surface water sites were mainly located downstream of wastewater effluent discharges and considered to be contaminated by human fecal pollution (Stewart-Pullaro et al., 2006).

Sewage-polluted tropical river waters and animal fecal samples were examined for male-specific RNA, male-specific DNA and somatic coliphages. Male-specific DNA and FRNA occurred at similar quantities per gram feces or per ml of water at 7% and 6.5%, respectively. F-specific phages were isolated by plaque assay from 50% of the river samples and 4.4% from the animal fecal material. The study concluded the presence of FRNA phages in the tropical waters of Klang Valley, Malaysia, and proposed FRNA as a tool for monitoring fecal-polluted waters in their country (Yee et al., 2006).

A small-scale study using male-specific, somatic phages and F+ *E. coli* as tracers of sewage, of monthly fluctuations and of various populations were investigated in a small Israeli community, Kibbutz Yagur, near Haifa, Israel. The well-defined sewage lines collect from daycare centers, residential, factory, dairy, greenhouse, dining hall, laundry room and

clinic with manholes specific to each location. Phages were not isolated from the manholes that served the laundry area and were found to occur in very low numbers (<10 PFU/ml) from the elderly home, the day-care center that used pampers, the greenhouse and the factory. The dairy farm was prevalent in somatic phages (10^2 - 10^3 PFU/ml) but had low counts of male-specific phages (1- 10 PFU/ml). Direct fecal material from newborn infants contained male-specific phages at concentrations of 10 - 10^5 PFU/g feces with one child excreting phages for almost 8 months. Absence or presence of F+ *E. coli* in the same water sample(s) correlated with low or high male-specific phage counts. Throughout the year-long study, there were higher numbers and more positive samples of male-specific phages than somatic phages with approximately 96-98.5% of typed phages being FRNA (Gino et al., 2007). Phages were not genogrouped or typed in this study.

FRNA Host Specificity

E. coli K-12 strain was studied in terms its ability or inability to transfer a sex factor, termed F. If cells transferred F to other cells by means of chromosomal markers, this was termed Hfr strains. If the F factor was transferred independent of the chromosome but through an extra-chromosomal state, or plasmid, the strains were known as F+ (Clark, 1963). Electron microscopy shed light onto the aggregate nature of FRNA phages adsorbed to the host cell's fimbriae, the F pili (Bradley, 1964) (Fig 4.1). The conjugative pili serve to transfer genetic information by horizontal gene transfer in gram-negative bacteria. Both Hfr and F+ strains are derived from *E. coli* K-12 (Havelaar and Hogeboom, 1984).

The male-specific DNA and RNA coliphages adsorb to these conjugative pili (Daehnel et al., 2005), a fertility (F+) sex-pili, coded in *E. coli* by the F-plasmid (Paranchych, 1975). DNA and RNA coliphages bind to the F-pili in different manners.

DNA phages bind to the tip of the F-pili and RNA phages attach to the sides (Daehnel et al., 2005).

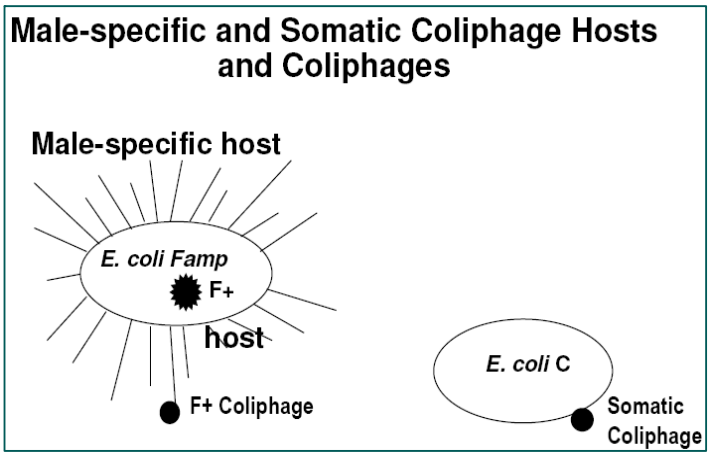
The Furuse studies enumerated FRNA phages by incorporating male strains of *E. coli* (F+, F' or Hfr) into the media plates followed by RNase treatment (Furuse, 1987). *Salmonella typhimurium* strain WG 49 that express F+ by the presence of an F-plasmid have also been used as a selective host to isolate male-specific phages (Havelaar et al., 1984). *S. typhimurium* detected somatic and FDNA phages, however, the host selected 90-95% FRNA phages (Sundram et al., 2006). *E. coli* strain F_{amp} (ATCC 700891) is commonly used in host-selection procedures for F+ coliphages.

Phage replication is restricted to environments such as the intestinal tract of warm-blooded animals as the sex-pili are only expressed at temperatures greater than 30 ^{o} C (Grabow, 2001). Interesting to note, Zinder (1963) and Bradley (1964) theorized that in order for male-specific phages to be plentiful in nature, then subsequently, the *E. coli* male strains must either be equally prevalent in nature or they are not the natural environmental host. However, bacteria from animal intestines harbor hosts with pili at 10^4 CFU/ml in wastewater, thereby possibly allowing phage replication (Long et al., 2005) in municipal wastewater systems.

FRNA phages are more abundant in sewage waters (10^3 - 10^4 PFU/ml) than in human and animal feces (up to 10^3 per gram feces) (Gerba, 2006). To examine this phenomenon, FRNA strain GA was inoculated into pasteurized sewage, seawater, and/or river water. Infectious strain GA underwent multiplication at 20^{o} C but only when host bacterium was cultivated at 37^{o} C (Havelaar and Pot-Hogeboom, 1988). Due to temperature restrictions of

the host-pili expression, it seems plausible that FRNA phage environmental replication would be restricted to > 30 ^{o} C. Optimal phage growth temperatures were found to be 30 ^{o} C for genogroup II and 37 ^{o} C for genogroups I, III and IV when three strains per genogroup were evaluated (Furuse, 1987). Other studies also reported different survival temperatures in water temperatures $\geq$ 15 ^{o} C (Sobsey, ftp.sccwrp.org). In addition, phage replication requires a host density of approximately 10^4 bacteria/ml to be successful (Goyal et al., 1987). The need for concomitant events of optimal temperatures and the presence of an F+ host in log phase would restrict environmental replication of FRNA coliphages.

Figure 4.1 Somatic and male-specific coliphage hosts.



Additional Potential Viral Indicators of Fecal Pollution

Bacteriophages are ubiquitous in nature and as a whole, are not suitable as an environmental water quality fecal indicator. However, unique bacteriophages are associated with sewage. Various phages have been proposed as viral indicators. Somatic coliphages (Fig 4.1), FRNA coliphages, FDNA coliphages and the *Bacteriodes* phage have been recommended. Advantages and limitations of each indicator will be addressed.

Somatic coliphages are present among four bacteriophage taxonomic groups, *Myoviridae*, *Styloviridae*, *Podoviridae* and *Microviridae*, hence the morphologic heterogeneity. Three families of somatic phages contain dsDNA and one family, *Microviridae*, contain ssDNA. The somatic phages are non-enveloped and are present in sewage, often in high abundance in untreated sewage, ranging from 10^4 - 10^5 PFU/ml (Gerba, 2006). These coliphages have been detected in humans, chickens, pigs, other animals (Gerba, 2006) and are easily cultured using *E. coli* strain CN-13 (ATCC 700609). The concentrations of somatic phages are similar to FRNA, 10^2 - 10^4 infectious units per liter (Sobsey, ftp.sccwrp.org) The phage adsorb to the *E. coli* host and other enterobacteria species via the cell wall with a basic receptor site. A limitation of somatic phages is the possibility that the bacteria host origin, especially from environmental reservoirs, may not have originated from a fecal source. Some somatic phages may replicate in environmental waters in the absence of fecal pollution (Ashbolt et al., 2001; Gerba, 2006). If the presence of somatic phages were unrelated to fecal contamination, then it would not serve to predict human health risk (Leclerc et al., 2000).

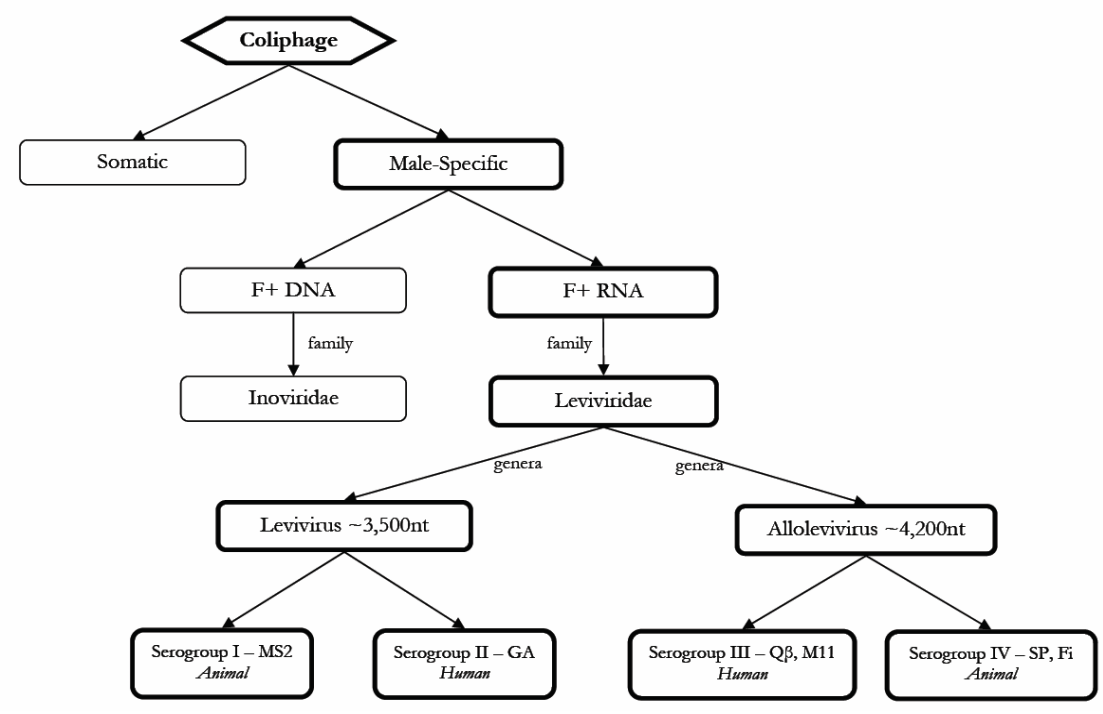
Inoviridae, a male-specific circular, single-stranded family of DNA viruses (Fig 4.2)

are filamentous, non-enveloped with a genome size of 6 kb to 9 kb and a virion length of approximately 700 nm to 2000 nm (www.virustaxonomyonline.com). Male-specific DNA coliphages are not morphologically similar to human enteric viruses and their ecology has not been extensively studied (Leclerc et al., 2000). Contradictions in the literature exist on whether FDNA or FRNA are more abundant in nature. Leclerc et al., (2000) noted FDNA were less abundant when compared to FRNA phages. Different distributions of FDNA and FRNA have been reported ranging from 52% FRNA and 48% FDNA (Vinjé et al., 2004) to 77% FDNA and 23% FRNA (Long et al., 2005). Cole et al., (2003) found higher numbers of FDNA in bovine wastes. Using Reverse-Line Blot hybridization and reverse-transcription-PCR followed by sequence analyses, one study found no link between FDNA type and source (Vinjé et al., 2004). However, a second study observed FDNA strain M13 as the dominant phage in septic system samples (Long et al., 2005). Thus, a definitive association between FDNA and human fecal pollution has not yet been established.

Certain phages belonging to the *Siphoviridae* family are approx 60 nm virion size, contain dsDNA with long non-contractile tails and infect via cell wall attachment to the host bacteria, strain *Bacteroides fragilis.* Host strain *B. fragilis* HSP 40 selects for a human-specific phage. The *B. fragilis* human-specific phage has been isolated in sewage, human fecal samples, polluted groundwater, seawater, river water and sediments but not detected in animal feces (Bitton, 2005; Cornax et al., 1990; Tartera and Jofre, 1987). Unlike other viral or bacterial indicators, the *Bacteroides* phage is considered to be an exclusive viral indicator of human fecal pollution and unlikely to replicate in the environment as the host is a strict anaerobe (Tartera et al., 1989). Human-specific phages can be rapidly detected with molecular primers without direct historical cultivation since animal vs human specific probes

are available (Bernhard et al., 2003). A limiting factor lies in the observations of a low isolation frequency, 0-15%, of *B. fragilis* phages were isolated from human feces (Gantzer et al., 2002). *B. fragilis* host are anaerobic and die-off rapidly under ambient water environmental conditions. As with other viruses, molecular detection only reveals the nucleic acid presence or persistence of the phage, not necessarily the infective or biologically active form. Another *Bacteroides* host, strain RYC2056, is more abundant but not human specific (Ashbolt et al., 2001). *B. fragilis* strain RYC2056 occurs in domestic sewage and was used to select phage from 30% of swine and 28% of human fecal samples (Puig et al., 1999).

Figure 4.2. Schematic of male-specific coliphage classification.



Coliphage and Enteric Virus Inactivation and Survival

As sewage and its accompanying microbial population are dispersed from point or non-point sources, microbes entering marine and freshwaters are subjected to the surrounding ecological conditions. Water-quality parameters such as pH, hardness, salinity, degree and intensity of sunlight exposure and temperature (climate) strongly influence microbe survival. Survival differs among microbial types, even from virus to virus. Virus types, enveloped or non-enveloped, vary greatly in environmental survival, as enveloped viruses tend to survive poorly outside of their host compared to the survival of non-enveloped viruses. F+ phages and enteric virus groups such as adenovirus, enterovirus, norovirus (calicivirus) and rotavirus are all non-enveloped viruses.

Survival was evaluated for the indicator organisms *E. coli, C. perfringens*, fecal coliforms and F+ phages in estuarine conditions. A series of parameters were selected such as temperature, salinity, dissolved oxygen, solar radiation, season and geographic location to study inactivation rates. The highest decay rates were influenced by sunlight and/or temperature. F+ phages exhibited the least decay (83%) whereas fecal coliforms had the highest decline at 99%. Phages and *C. perfringens* were least affected by temperature. This study provides corroborative evidence that F+ phage inactivation rates differ from bacterial indicators (Burkhardt III et al., 2000).

Representative prototype and genotyped field isolated strains of FRNA were spiked at concentrations of 10^5 - 10^7 infectious units into FRNA-free, untreated surface waters (freshwater). Samples were incubated at 25 ^{o} C in the dark and titered over time to determine phage decline or inactivation. Aliquots were removed for FRNA quantitation at day 4, 7, 13 and 36. Initial titers were determined by double-agar overlay plaque assay. The last sample taken, day 36, was enriched and assayed with the two-step enrichment procedure.

At day 4, rates of inactivation for prototype strains were as follows: genogroup I MS2 4.7 $\log_{10}$, genogroup II GA 2.6 $\log_{10}$, genogroup III QB 1.7 $\log_{10}$ and genogroup IV FI 2.7 $\log_{10}$ and SP (inactivation log data not provided). Prototype strain SP had the poorest survival and was not detectable after 4 days. Neither strain SP or QB could be detected by the two-step enrichment at day 36. Strains MS2, GA and FI were detected by enrichment at day 36. GA had the longest survival duration and was detectable by DAL throughout the experiment (day 36) (Brion et al., 2002). Survival of genotyped environmental isolates for FRNA coliphages was compared to their respective prototype strain. Of four, group I field strains, two survived similar to MS2 whereas the remaining two isolates were detectable by DAL to the end of the experiment, day 36. In group II, GA and one field II strain survived the longest (day 36) and the remaining three, group II field isolates declined by day 15. Type III isolates were below detection limits by day 13. Survivability was strain-specific and not necessarily influenced by genotype or prototype vs field isolates in freshwater systems (Brion et al., 2002).

A non-water matrix was used to compare the inactivation of MS2 to other pathogenic viruses, norovirus and poliovirus type 1, and *E. coli.* Different microcosms were filled with representative soils, i.e., organic muck, clay and sand. For one month, viruses seeded in a groundwater matrix at approx 10^6 were dosed into the soil microcosm twice/week. The following month, the columns were dosed with a simulation of rainwater. Microcosm effluent was drained and viral strains were detected by RT-PCR or infectivity assays. With the exception of *E coli*, the viruses were detected and shown to pass through the column. The results suggested that *E. coli* was not a reliable viral indicator as the bacteria did not pass through the columns in the manner as the virus and ultimately, environmental bacterial

transport would not mimic virus transport. None of the tested indicators passed through the clay column (Meschke and Sobsey, 2003).

Freshwater, specifically lake waters, were used in a microcosm format to evaluate survival rates on the four FRNA genogroups for 110 days. Strains MS2, two environmental group I, a group II, a group III and two group IV phages were inoculated (2 x 10^4 PFU/ml) into the microcosms. Incubation temperatures were 4 and 20 ^{o} C. Survival time was greater at lower temperature (4 ^{o} C). Group IV had the fastest decay rate (5 $\log_{10}$ within 10 days, 4 ^{o} C) followed by group III (5 $\log_{10}$ at 3 weeks, 4 ^{o} C), where these isolates reached the limits of detection at 10 days and 3 weeks, respectively. Rates of inactivation for groups I and II at 110 days, 4 ^{o} C were 1 $\log_{10}$ and 3 $\log_{10}$, respectively. The more persistent genogroups, groups I and II, were detected at 110 days (Long and Sobsey, 2004).

Prototype strains MS2, GA, Qβ and FI were spiked into seawater at concentrations of approximately 10^6 PFU/ml and incubated, in the dark, at 21-23 ^{o} C. One ml aliquots were removed daily and titers observed by double-agar overlay (DAL) and real-time PCR, concurrently. Within 7 days, all four subgroups in seawater were no longer detected by DAL. In contrast, the real-time PCR detected the four groups after 20 days (Kirs and Smith, 2007). PCR detection of RNA, as demonstrated in these experiments, does not predict or indicate the presence of infectious phage but only the presence of RNA.

Viruses, including phage MS2, were tested as a molecular model for human norovirus. Spiked surface and groundwater samples were assayed for viral nucleic acid presence at 25 and 4 ^{o} C using QRT-PCR. In some strains, cell-culture infectivity and the DAL method for MS2 were used as concurrent methods of detection. MS2 nucleic acid detection was not effected by water source. Infectivity reduction rates were observed to

significantly increase at 25 ^{o} C compared to 4 ^{o} C. Infectivity reduction rates were not observed at 4 ^{o} C (Bae and Schwab, 2008).

FRNA inactivation rates are influenced by sunlight intensity, fresh vs saltwater matrices and soil composition. For example, increasing salinity and temperature demonstrated faster rates of inactivation in FRNA strains. In contrast, lower temperatures (4 ^{o} C) and freshwater environments decreased inactivation rates on FRNA coliphages.

FRNA Coliphage Detection Methods - Plaque Assays

To quantify and accurately detect FRNA phages, initial purification methods are imperative. Depending upon the application, detection methods include Single Agar Overlay,

Double-Agar Overlay, enrichment, Most Probable Number, RNase sensitivity, serotyping and genotyping.

To prepare a Single Agar Overlay (SAL) for enumeration of FRNA, collect 250 ml environmental water sample and place on ice up to 6 hrs. Beforehand, begin a log-phase of the *E. coli* host F_{AMP} (ATCC # 700891) by adding a 1ml inoculum from an overnight (O/N) culture into 50 ml Trypticase Soy Broth. Gently shake at 37 ^{o} C for 4 hr. Host culture will be in log-phase at 4 hr. Divide the environmental sample by dispensing two aliquots of 100 ml into sterile bottles. To the 2X TSA (trypticase soy agar) add 1 ml of 100 X streptomycin/ampicillin, 0.5 ml sterile 4M $MgCl_2$. Add 10 ml log-phase F+ host to the 100 ml sample. Quickly combine the sample/*E. coli* host with the 2X TSA/antibiotics flask. Gently swirl and allow host to contact sample for at least 3 minutes before pouring plates. Pour the agar/sample mix into a series of 20mm petri dishes at approx 20 ml/petri dish. Allow to harden and dry. Invert and incubate 36 ^{o} C $\pm$ 1 ^{o} C. Count plaques per plate as

PFU/100 ml. Questionable plaques can be verified by spot plate assay (EPA Method 1602, 2001).

A presence/absence method involves a two-step enrichment. Using a sterile bottle, divide the 1 L environmental water sample into two, 500 ml aliquots. To each 500 ml, add the following: 50 ml of 10X cold TSB per L (media is cold to diminish phage growth), 12.5 ml 4M $MgCl_2$, 10 ml of 100 X streptomycin/ampicillin and 5 ml log-phase F+ host. Cap and invert to mix. Incubate overnight (16-24 hr) at 36 ^{o} C $\pm$ 1 ^{o} C. The following day, spot 10 ul from each enrichment onto a grid spot plate (TSA) containing the appropriate host and incubate overnight. Score as positive or negative lysis zones in spots (EPA Method 1601).

An MPN (Most Probable Number) can be estimated from a 1L environmental water sample. Enrich the 1L sample with 2X TSB, antibiotics, 4M $MgCl_2$ and *E. coli* host as stated above. Once the reagents are added, quickly aliquot the enrichment into the following amounts in triplicate (A,B,C): (i) 300 ml A, 300 ml B, 300 ml C (ii) 30 ml A, 30 ml B, 30 ml C and (iii) 3 ml A, 3 ml B, 3 ml C. This dilution series is for a 3x3 MPN matrix Incubate bottles/tubes at 37 ^{o} C O/N. Remove 12 ml from each enrichment of 300 and 30 ml (A,B,C) except the 3 ml. Centrifuge all samples for 10 minutes at approx 5000 rpm. Transfer 10 ml of each supernatant into clean tubes; transfer most of the 3 ml supernatant into a clean tube. Briefly vortex the MPN supernatant and apply 10 ul from the appropriate supernatant onto the respective labeled spot (spot plate). Allow the spots to dry approx 20-30 min. Invert and incubate at 37 ^{o} C O/N. Score as presumptive positive if a clear zone of lysis is visible in the spot and calculate the MPN from the number of positive (lysis zone) and negative (no lysis zone) spots. If necessary, verify the lysis zone by isolation of material from it and a repeated spot plate of this material.

Applications to detect somatic phages are the same as those for male-specific except the selection host is one that is F-minus (F-) such as *E. coli* strain CN-13 (ATCC 700609). For this host a 1% stock of antibiotic nalidixic acid is used in the media instead of strep/amp as used for *E. coli* Famp (EPA Method 1601, 2001). *E. coli* host C-3000 (ATCC 15597) selects for both F+ and somatic phages but findings from field studies reveal strain C-3000 recovers lower numbers of somatics and male-specific phages when compared to the sum of phages recovered by their respective host-specific strain. In a recent study, *E. coli* strain CB390 was effective at recovering the sum of both phage types (Guzman et al., 2008).

Coliphage antisera are obtained by inoculating experimental animals to elicit an immune response that results in polyclonal antibodies against the desired phage strain. For serotyping, isolated plaques are spotted (10 ul) onto an agar plate containing a specific antisera, either to MS2, GA, Qβ, SP or FI. The agar plate contains the antisera and log phase host. Once the plaque is applied, the plate is incubated O/N at 37 ^{o} C. Phage growth suppression on one of the plates, which contains homologous neutralizing antibodies to the phage group, is scored as positive. Serotyping is not 100% reliable and can produce ambiguous results (Beekwilder et al., 1996). For example, two isolates in the UNC collection were first serotyped to group II but later sequenced and genotyped to group I.

RNase sensitivity is determined by re-plating phage isolates on plates containing or lacking Ribonuclease A. On the plates with RNase, FRNA phages do not form lysis zones, however, FDNA phages will form zones of lysis.

FRNA Coliphage Detection Methods - Genotyping

Identification of male-specific phage by host selection does not discriminate between

FDNA and FRNA phages. An FRNA assay applicable for microbial source-tracking must first and foremost precisely identify FRNA. Genogroup-targeted methods solve two considerations, (i) FRNA selection and (ii) genotype-associated source tracking.

To distinguish the individual FRNA groups I, II, III, IV and a combination of groups I and II, III and IV, a molecular hybridization assay was developed. Genogroup-specific hybridization probes were designed using multiple alignment software to strains MS2, GA, Qβ, and SP. Environmental isolates of FRNA phage as lysis zones on agar plates were adsorbed to a membrane, denatured to unfold RNA secondary structures and linked to the membrane using UV light. Digoxigenin (DIG-dUTP) labeled probes were incubated overnight with the fixed membranes and visualized with alkaline phosphatase-conjugated anti-DIG antibody. Assay development involved optimization of various hybridization solutions, selection of the most efficient membrane and denaturation solutions. To test the hybridization assay, 203 FRNA field isolates from sewage, oysters, surface waters and feces were first grouped by serotyping. Serotyping and hybridization classified 79 and 109 isolates, respectively, from surface waters, adult and piglet swine feces, treated and untreated sewage and oysters as genogroup II. Almost half of the piglet isolates, 12 out of 26, were type IV. Isolates from chickens were grouped almost equally into I and IV. Serotyping cross-reactivity was observed for 37 isolates neutralized by anti-sera GA and partial neutralization to anti-sera MS2. However, these 37 serotyped strains were eventually genotyped by hybridization to be in group II. Similar classification results were obtained by both serotyping and hybridization genotyping (Hsu et al., 1995).

A hybridization assay (Beekwilder et al., 1996) was developed similar to Hsu et al., (1995), but oligonucleotide probes targeted different regions of the genome. Probe to group I

was designed to MS2 nucleotide region 1248 (Hsu) and 1260 (Beekwilder), group II was designed to GA region 431 (Hsu) and 2100 (Beekwilder), group III was designed to Qβ region 27 (Hsu) and 660 (Beekwilder) and group IV probe was designed to SP region 35 (Hsu) and 40 (Beekwilder). Probes were designed to each genogroup by Beekwilder and colleagues by alignment of 3-5 strains/group. If the completed sequenced genome was not available, they proceeded to sequence partial regions to provide adequate sequence representation for each genogroup. In addition, a probe A for *Levivirus* (groups I and II) and probe B for *Allolevivirus* (groups III and IV) was developed. To validate the hybridization probes, a combination of "blinded" but previously serotyped samples and field samples were analyzed. Approximately 78% of the samples were correctly identified by both the genogroup and genus-specific probes. Isolates collected from human impacted areas in the form of hospital waste and domestic wastewater identified 1 group I, 1 group II, 4 group III and 1 group IV. The two human feces samples had ambiguous classifications in that both samples were weakly positive for group I and positive for group IV but hybridized to both genus-specific probes, *Levivirus* and *Allolevivirus*. Twenty phages isolated from animal sources hybridized to groups I or IV in all cases except two. Of those two, one isolate was positive to groups I and II, and the second isolate (porcine slaughterhouse) was positive to group III (Beekwilder et al., 1996). These hybridization studies lend supporting evidence to the trends that genogroups II and III are associated with human waste and groups I and IV occur more often in animal sources.

Reverse-line blot is a nucleic hybridization-based assay designed to genotype both FRNA and FDNA coliphages. Using Clustal W 1.4 software, alignment of complete or partial sequences available in GenBank was the basis for primer and probe design. Six

cluster-specific oligonucleotide probes for hybridization were designed to FRNA strains, 1 group I probe for MS2-like, 1 probe for GA-like, 2 group III probes, Qβ and M11, and 2 group IV probes, SP and FI-like. Three cluster-specific FDNA probes were also designed, M13, fd and CH and one generic FDNA consensus-specific probe, termed "con." In order to initially divide FRNA from FDNA isolates, a generic RT-PCR assay was performed. Broadly reactive primer sets yield different PCR amplicon sizes for *Allolevivirus* vs. *Levivirus*. *Allolevivirus* was amplified by primers MJV82 forward, JV41 reverse and *Levivirus* was detected by primers MJV82 forward and JV81 reverse. Primer pair SL2 forward and SL3 reverse amplified FDNA phages. Assay validation began by collecting a total of 557 environmental samples. Phages were isolated by single or double-agar overlay plating, followed by RNase sensitivity testing and serotyped by anti-sera neutralization or genotyped by hybridization (Hsu et al., 1995). RT-PCR identified 100% of the FRNA and FDNA strains. Identified strains were used to validate the reverse-line blot hybridization (RLB), resulting in 98% agreement of the FRNA strains and 100% confirmation of the FDNA phages (Vinjé et al., 2004).

When comparing the basic hybridization assays, the RLB involves a two-step process since the RT-PCR step occurs prior to hybridization. Nonetheless, RLB typing had a higher predictability, 100% FDNA, 98% FRNA, when compared to 38% by serotyping (Hsu et al., 1995), 54% by genotyping (Hsu et al., 1995) and 78% by genotyping (Beekwilder et al., 1996).

Prior to 2007, 10 full-length or nearly full-length FRNA phage genomes were available at NCBI GenBank. In 2007, Kirs sequenced strain FI (Kirs and Smith, 2007), for a total of 11 FRNA genomes. Studies describing molecular detection of FRNA phages base their primer design on these 10 or 11 sequences and partial sequences available in GenBank.

A universal forward and reverse primer set was designed (Kelly Reynolds, Univ of AZ) to a consensus sequence in the replicase gene to detect all FRNA strains. A two-step reverse transcription polymerase chain reaction (RT-PCR) was performed. Prior to RT-PCR, environmental samples were column filtered to remove inhibitors. RT-PCR sensitivity was improved when the samples were column filtered and detection limits were 0.10 PFU of laboratory control MS2. RT-PCR amplified FRNA from 2 samples that were plaque-negative by soft agar overlay and was in agreement with the plaque-positive overlay methods (Rose et al., 1997). According to the authors, the primers amplified all four FRNA coliphage groups, but specific strains were not mentioned.

Phage MS2 is routinely used as a surrogate for pathogenic and environmental studies. Five sets of real-time primers and probes were designed to strain MS2. Two sets targeted the assembly gene, 1 for the coat region, 1 targeted the lysis gene and 1 primer and probe set targeted the replicase gene. Cross-reactivity of the real-time PCR was tested against non-targeted organisms, specifically pathogenic bacteria. MS2 primer sets did not cross-react with bacteria (O'Connell et al., 2006). Detection or cross-reactivity to other FRNA phages was not discussed.

Two independent real-time assays to detect each FRNA genogroup were developed for microbial source-tracking. Primer and probe sets were designed to the limited genomic NCBI GenBank database, by aligning 2 to 3 strains/genogroup and partial sequences. Purified RNA and a two-step QRT-PCR assay were common to both investigations. Kirs and Smith (2007) used a multiplexed assay whereas Ogorzaly and Gantzer (2006) analyzed each genogroup in separate vials. QRT-PCR primer and probe specificity were evaluated in the reaction containing an RNA cocktail of the strains MS2, GA, Qβ, SP (Ogorzaly and

Gentzer, 2006) and MS2, GA, Qβ, SP and FI (Kirs and Smith, 2007). Although the authors state their assay was template specific and lacked cross-hybridization, assay validation was limited to four or five FRNA strains. Raw sewage (Ogorzaly and Gentzer, 2006) and raw sewage and chicken litter (Kirs and Smith, 2007) samples were collected for QRT-PCR field validation. Ogorzaly and Gentzer detected 100% of groups I, II and 85% group III by QRT-PCR of double-agar overlay plaques. Group IV was not detected. Twenty plaques from sewage and 20 plaques from chicken litter were isolated, RNA purified and subjected to QRT-PCR (Kirs and Smith, 2007). Sewage isolates were group III and chicken stool isolates were group IV. Three genogroup-specific primer pairs were developed for a two-step RT-PCR assay and validated with strains MS2, GA and SP. Primer design for genogroups I, II and IV were based on GenBank genomic sequences for MS2, GA and SP, respectively. Phages isolated and enumerated from individual septic systems, poultry farm, municipal sewage and a background site were enumerated by SAL after filtration of the water sample. An MPN was also conducted. Purified phage RNA, from October 2004 environmental samples, tested positive for groups I and IV with RT-PCR and was in agreement with plaque-positive samples. The septic system tested negative with RT-PCR but positive by SAL and MPN. In January, 2005, 12 positive SAL and MPN samples tested positive for group I by RT-PCR. The May, 2005, sampling season had 3 out of 7 samples positive by SAL and MPN detection. In those 3 samples, RT-PCR identified group II upstream from the wastewater treatment plant and group I from the poultry farm-area of the lake. In contrast, SAL-positive samples downstream from the sewage treatment plant were negative by RT-PCR. Similar results were obtained in the August, 2005, sampling season as 4 samples were positive by SAL and MPN. Of those 4 samples, RT-PCR identified group II in the upstream

wastewater treatment plant sample but the 3 remaining samples taken downstream of the sewage plant, upstream of a different sewage plant and lake waters near the poultry farm were negative by RT-PCR. The authors concluded FRNA indicators were not useful to distinguish between human and non-human sources and suggested the presence of either somatic or FDNA phages in their positive samples (Dryden et al., 2006). Limitations to their study design are as follows. Primers were designed to three individual strains and not an alignment of multiple strains per genogroup. These primers would not detect all four genogroups or environmental strains as reflected in the results. A group III primer set was not designed and yet, the authors concluded FRNA could not distinguish between human and non-human sources. SAL and MPN host selection was *E. coli* strain C-3000 (ATCC 15597), a host that selects for both somatic and male-specific phages.

An antibody-based agglutination assay, termed "latex agglutination", was developed to rapidly detect, <24 hr, genogroups of FRNA coliphages. Environmental strains were collected from bird droppings, shellfish and water bodies from diverse geographical locations across the USA. A two-step enrichment protocol was modified to a culture time of 180 min based on preliminary sampling and phage measurement at 0, 30, 60, 90, 120, 180 and 360 min on TSA plates containing *E. coli* host F_{AMP} . Male-specific plaques were confirmed as FDNA or FRNA by RNase sensitivity, FRNA genera were determined by broadly reactive RT-PCR and preliminary genogrouping by RLB (Vinjé et al., 2004). Serotyping and sometimes even RLB genotyping, at times, yield ambiguous results. To address this ambiguity, an RT-PCR assay to the capsid region with primers DL10 forward and DL11 reverse and sequencing of the PCR amplicon was conducted. To develop the coliphage latex

agglutination and typing (CLAT) assay, polyclonal antibodies were first generated against MS2, GA, Qβ, SP and FI. Phage strain-specific antisera was bound to polystyrene particles and used in the agglutination step of the CLAT. Coliphage enriched samples showed agglutination within 30-60 sec when 2.5 ul enriched phage cultures were mixed with 2.5 ul strain-specific antibody particle on an agglutination card. Out of 192 FRNA field isolates, CLAT correctly identified 185 (96%) and RLB identified 92%. When the two methods were compared, some ambiguity was noted in that CLAT identified more group II and RLB identified more groups III and IV. Twenty-four strains yielded inconsistent typing. Sequencing of the capsid region and phylogenetic clustering of these 24 strains yielded 19 group I and 5 group II. CLAT clustered 17 of the 19 group I strains as I and II, but matched

100% of the 5 sequenced group II isolates. A total of 164 FDNA isolates were also identified by CLAT at a rate of 97.7% (Love and Sobsey, 2007).

FRNA Relationship to Enteric Viruses and as Predictors of Health Risk

An important attribute of an ideal indicator is the relationship between the indicator density in polluted waters to human-health risks. Few epidemiology studies exists correlating health risks with F+ coliphage densities. A European cohort study (Lee et al., 1997) showed a statistically significant relative risk (RR) with increased F+ coliphage exposures. However, a threshold value was not extrapolated and the slight increased RR (2.6-2.8) is minimal compared to the large coliphage density range of 26 - 32 and 69-308 PFU/10 ml, respectively.

A recent California beach study suggested an association between F+ coliphage densities (both FRNA and FDNA combined) and gastrointestinal illness rates, nausea, cough

and fever. The authors noted that when F+ phages were detected, a low number of beach-goers were exposed (Colford et al., 2007).

Meta-analysis of epidemiological studies revealed no correlation between bacteriophage in marine waters to GI illness, whereas the freshwater studies reported an elevated GI risk with elevated bacteriophage exposure (Wade et el., 2003). The meta-analysis did not clarify which type of bacteriophage showed associations with human health risks.

Six wastewater treatment facilities from FL, AZ and CA producing and distributing reclaimed water were monitored for indicator and pathogen load. Although reclaimed water is routinely assessed for fecal or total coliforms, the degree of microbial indicator and pathogen removal has not been evaluated. Results using male-specific coliphages did not show a correlation with enteric viral load. However, the coliphage predicted an absence of enteric viruses at levels less than 10 coliphage/100 ml (Rose et al., 2004).

Adenovirus presence was statistically correlated to FRNA ($r = 0.99$) in a brackish (salinity from 9-34 ppt) but no correlation to FDNA phages were observed in coastal waters (Jiang et al., 2001).

Future Applications

Molecular detection methods, i.e., PCR, real-time PCR and microarrays, allow a more timely assessment of the microbial quality of recreational waters. A prospective study at two Great Lakes beaches using real-time PCR detection found that enterococci was statistically associated with increased gastro-intestinal (GI) illness at both beach sites. A strong positive trend was noted with the presence of *Bacteroides* at one of the beaches (Wade et al., 2006).

Microarrays have been constructed with genomic DNA purified from raw wastewaters (Lee and el., 2006) and 16S rRNA and cpn60 genes extracted from several specific pathogens (Maynard et al., 2005). Primers and oligonucleotide microarray probes were designed from sequences derived from specific pathogenic bacteria strains. Validation of the microarrays generated a positive hybridization signal in raw sewage samples for the *E .coli* gene *uidA* (Lee et al., 2006).

Commercially-available field PCR instruments could potentially be applicable to molecular detection of FRNA genogroups. For example, a hand-held fluorogenic real-time PCR instrument, the Advanced Nucleic Acid Analyzer (ANAA), was used to detect bacteria spores and MS2 virions. The microbes were analyzed using a micro-chip in the ANAA and a positive signal was viewed within 18-26 min (Belgrader et al., 1998).

Portable real-time thermocyclers (Cepheid, Inc) and nucleic acid sequence based amplification (NASBA) instruments (BioMereux, Inc) can be transported to provide on-the-spot analysis of environmental samples. NASBA, unlike PCR, does not require a thermocycler. NASBA relies on an isothermal process (41 ^{o} C), three RNA-associated enzymes and a molecular beacon. Single-stranded RNA is generated in a single-tube, emitting fluorescence by the molecular beacon upon hybridization with ssRNA and a target-specific oligonucleotide. NASBA technology targets RNA viruses and would be applicable as a portable instrument to detect FRNA coliphages.

Summary and Conclusion

Microbial source-tracking is defined as a group of analytical protocols, typically microbial applications, which are used to ascertain or regulate the source of fecal input into a water body. Use of a potential microbial source-tracking “toolbox” is a complex process

whereby decisions must be made to select and validate the selection of such tools. For example, (1) define the problem (nearby source discharge or runoff, history of monitoring data, weather patterns, hydrology, etc), (2) formulate objectives as to the suspected source and category (human vs. non-human), (3) presence/absence vs. quantification of loading values (4) consider if the application linked to regulatory values for microbial indicator concentration or if fecal presence or absence is an acceptable result for decision making, (5) consider if fecal presence is linked to public health risk, (6) consider if there are legal ramifications, (7) select the most appropriate source-tracking protocol (8) define a sampling strategy or study design, (9) define a method of data collection and quality control, (10) define data analyses and (11) have a plan for data interpretation (Stoeckel, 2005).

Currently, the toolbox consists of DNA fingerprinting, antibiotic resistance, ribotyping, pulse-field electrophoresis and carbon utilization profiles of *E. coli* and/or enterococci (library-dependent), host-specific *Bacteroides* and *Prevotella* bacteria markers (library-independent) and serotyping or genotyping of FRNA or FDNA coliphages (library-independent). The "library" per se, is a collection of bacterial or viral isolates from which the source of collection is known as well as the fingerprint, marker, genotype or serotype. One aspect of method selection depends upon the analytical question, library-dependent (epidemiological matching or clustering) vs. library-independent (source could be traced in any water body type or geographical location).

FRNA source-specificity displayed by the four genogroups render FRNA phages applicable for microbial source-tracking. Groups I and IV are typically associated with animal wastes whereas groups II and III generally occur from human waste (Havelaar et al., 1986; Schaper et al., 2002). FRNA library collections began with known sources, i.e.,

domestic sewage/drainage or hospital waste sites, cattle, swine and poultry waste lagoons or litters, animal-specific fecal voids, and/or animal intestinal content and, following FRNA purification, these sources were further defined by serotyping or genotyping (Furuse et al., 1978; Miyake et al., 1971; Furuse et al., 1975; Miyake et al., 1973; Osawa et al., 1981; Schaper et al., 2002, Beekwilder et al., 1996; Hsu et al., 1995; Calci et al., 1998; Cole et al., 2003; Long et al., 2005; Sundram et al., 2006; Stewart-Pullaro et al., 2006; Kirs and Smith, 2007). Building upon trends in FRNA source specificity, investigators have isolated environmental phages, purified them and genotyped/serotyped as a strategy to classify fecal contamination (Dryden et al., 2006; Love and Sobsey, 2007; Stewart-Pullaro et al., 2006; Vinjé et al., 2004; Stewart et al., 2006; Yee et al., 2006; Sundram et al., 2006; Brion et al., 2002; Sobsey et al., 2006).

Types and trends of known FRNA host specificity were reported to indicate human source from sewage-collected FRNA as follows: genogroup I 38% , 15%, 12%, 11%, 2%; genogroup II 50%, 38% , 21% , 15%, 14%; genogroup III 100%, 73%, 58%, 57%, 50 %, 46%, 15%, 6%; and genogroup IV 1%, 5% IV and animal sources: genogroup I 98%, 70%, 51%, 50%, 50%, 32% 30%; genogroup II 15%, 5%; genogroup III 22%, 19%, 18%; and genogroup IV 100%, 50%, 50%, 46%, 6%, 3% (Aoi et al., 1972; Miyake et al., 1973; Brion et al., 2002; Miyake et al., 1971; Cole et al., 2003; Hsu et al., 1995; Kirs and Smith, 2007). Although the data from these studies is not comprehensive or exhaustive, a host-specific FRNA genogroup trend exists.

In this dissertation, a library-independent method to differentiate between sources of human and non-human fecal pollution was developed. Initially, nineteen FRNA strains were sequenced and compared to the eleven FRNA full-length sequences available in the National

Center for Biotechnology Information (NCBI) genetic database (GenBank) for a total of 30 FRNA strains. FRNA phages were collected from water, sewage, and various animals representative of diverse geographical locations (Table 7.1). The field-collected FRNA strains and prototype strains were represented by phages isolated from the United States, Japan, Europe, Brazil, Taiwan and/or Mexico (Table 7.1). FRNA sequences generated in this study tripled the genetic information currently available in the national genetic database for the *Leviviridae*. Based on the intensive sequencing effort, a robust, one-step reverse transcription polymerase chain reaction (RT-PCR) was designed to distinguish the four FRNA coliphage groups (I, II, III, IV). In the more immediate future this data can be applied to methods using FRNA coliphages as a fecal and viral indicator and as a source-tracking tool.

To conclusively apply the "FRNA group II, III human fecal pollution" association for recreational or other surface water regulatory purposes, one must caution that one or two or a few isolates does not confirm a human-impacted area. Perhaps a more plausible application would be to propose an absolute percentage or value of group II and III isolates from a pre-determined FRNA sample size detected from an area where wastewater discharges and/or other source delineations have been examined.

References

Ackermann HW. 2006. Classification of bacteriophages. In: *The Bacteriophages* by S T Abedon. R L Calendar, ed. Oxford University Press. pp:8-16.

American Water Works Association. 2004. Field testing of USEPA methods 1601 and 1602 for coliphage in groundwater. Prepared by MR Karim, MW LeChevallier, M Abaszadegan, A Alum, J Sobrinho and J Rosen. AWWA Research Foundation, Denver, CO.

Armon R and Y Kott. 1995. Distribution comparison between coliphages and phages of anaerobic bacteria (*Bacteroides fragilis*) in water sources, and their reliability as fecal pollution indicators in drinking water. Wat Sci Tech, 31(5):215-222.

Ashbolt NJ, WOK Grabow and M Snozzi. 2001. Indicators of microbial water quality. In: *World Health Organization, Water Quality: Guidelines, Standards and Health.* L Fewtrell and J Bartram, eds. IWA Publishing, London, UK. pp:289-316.

Aoi T, K Furuse, T Shiba and I Watanabe. 1972. Isolation and grouping of RNA phages III. A survey in Arabia, India and Thailand. J Keio Med Soc, 49:361-368.

Bae J and KJ Schwab. 2008. Evaluation of murine norovirus, feline calicivirus, poliovirus, and MS2 as surrogates for human norovirus in a model of viral persistence in surface water and groundwater. Appl Environ Microbiol, 74(2):477-484.

Beekwilder J, R Nieuwenhuizen, AH Havelaar and J van Duin. 1996. An oligonucleotide hybridization assay for the identification and enumeration of F-specific RNA phages in surface water. J Appl Bacteriol, 80:179-186.

Belgrader P, W Bennett, D Hadley, G Long, R Mariella, Jr, F Milanovich, S Nasarabadi, W Nelson, J Richards and P Stratton. 1998. Rapid pathogen detection using a microchip PCR array instrument. Clin Chem, 44(10:2191-2194.

Bernhard AE. T Goyard. , MT Simonich and KG Field. 2003. Application of a rapid method for identifying fecal pollution sources in a multi-use estuary. Wat Res 37(4):909-913.

Bitton, G. 2005. Microbial indicators of fecal contamination: application to microbial source tracking. Report submitted to the Florida Stormwater Association, Tallahassee, FL. June, 2005.

Borrego JJ, MA Morinigo, A de Vicente, R Cornax and P Romero. 1987. Coliphages as an indicator of faecal pollution in water. Its relationship with indicator and pathogenic microorganisms. Wat Res, 21(12):1473-1480.

Bradley DE. 1964. The structure of some bacteriophages associated with male strains of

Escherichia coli. J Gen Microbiol, 35:471-482.

Brion GM, JS Meschke and MD Sobsey. 2002. F-specific RNA coliphages: occurrence, types, and survival in natural waters. Water Res, 36:2419-2425.

Buchen-Osmond, C. (Ed.) 2004. Levivirus. In ICTVdB - *The Universal Virus Database,* version 3. ICTVdB Management, Columbia University. New York, NY.

Burkhardt III W, KR Calci, WD Watkins, SR Rippey and SJ Chirtel. 2000. Inactivation of indicator microorganisms in estuarine waters. Wat Res, 34(8):2207-2214.

Clark AJ. 1963. Genetic analysis of a "double male" strain of *Escherichia coli* K-12. Genetics 48:105-120.

Calci KR, W Burkhardt III, WD Watkins and SR Rippey. 1998. Occurrence of male-specific bacteriophage in feral and domestic animal wastes, human feces, and human-associated wastewaters. Appl Environ Microbiol, 64(12):5027-5029.

Cole, D, SC Long and MD Sobsey. 2003. Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. Appl Environ Microbiol, 69(11):6507-6514.

Colford JM, Jr., TJ Wade, KC Schiff, CC Wright, JF Griffith, SK Sandhu, S Burns, M Sobsey, G Lovelace and SB Weisberg. 2007. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology 18(1):27-35.

Cornax R, MA Morinigo, IG Paez, MA Munoz and JJ Borrego. 1990. Application of direct plaque assay for detection and enumeration of bacteriophages of *Bacteroides fragilis* from contaminated water samples. Appl Environ Microbiol, 56(10):3170-3173.

Daehnel K, R Harris, L Maddera, and P Silverman. 2005. Fluorescence assays for F-pili and their application. Microbiology, 151:3541-3548.

Dryden SK, B Ramaswami, Z Yuan, DE Giammar, LT Angenent. 2006. A rapid reverse transcription-PCR assay for F+ RNA coliphages to trace fecal pollution in Table Rock Lake on the Arkansas-Missouri border. Water Res, 40:3719-3724.

Duckworth DH. 1987. History and basic properties of bacterial viruses. In: *Phage Ecology*. SM Goyal, CP Gerba and G Britton, eds. John Wiley & Sons, Inc., NY. pp: 1-43.

Fiers W, R Contreras, F Duerinck, G Haegeman, D Iserentant, J Merregaert, W Min Jou, F Molemans, A Raeymaekers, A Van den Berghe, G Volckaert and M Ysebaert. 1976.

Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature, 260:500-507.

Furuse K. 1987. Distribution of coliphages in the environment: general considerations. In: *Phage Ecology*. SM Goyal, CP Gerba and G Britton, eds. John Wiley & Sons, Inc., NY. pp 87-124.

Furuse K, A Ando and I Watanabe. 1975. Isolation and grouping of RNA phages VII. A survey in Peru, Bolivia, Mexico, Kuwait, France, Australia and the United States of America. J Keio Med Soc, 52:353-361. (Japanese)

Furuse K, T Sakurai, A Hirashima, M Katsuki, A Ando, and I Watanabe. 1978. Distribution of ribonucleic acid coliphages in South and East Asia. Appl Environ Microbiol, 35(6):995-1002.

Gantzer C, J Henry and L Schwartzbrod. 2002. *Bacteroides fragilis* and *Escherichia coli* bacteriophages in human feces. Int J Hyg Environ Health, 205:325-328.

Gerba, CP. 2006. Bacteriophage as pollution indicators. In: *The Bacteriophages* by S T Abedon. R L Calendar, ed. Oxford University Press. pp: 695-701.

Gino E, J Starosvetsky and R Armon. 2007. Bacteriophage ecology in a small community sewer system related to their indicative role in sewage pollution of drinking water. Environ Microbiol, 9(10):2407-2416.

Guzman C, L Moce-Llivina, F Lucena and J Jofre. 2008. Evaluation of *Escherichia coli* host strain CB390 for simultaneous detection of somatic and F-specific coliphages. Appl Environ Microbiol, 74(2):531-534.

Goyal SM, CP Gerba and G Britton, eds. 1987. Phage Ecology. John Wiley and Sons, Inc. NYC, NY.

Havelaar AH, and WM Hogeboom. 1984. A method for the enumeration of male-specific bacteriophages in sewage. J Appl Bacteriol, 56:439-447.

Havelaar AH, WM Hogeboom, and R Pot. 1984. F-specific RNA bacteriophages in sewage: methodology and occurrence. Wat Sci Tech, 17:645-655.

Havelaar AH and WM Pot-Hogeboom. 1988. F-specific RNA-bacteriophages as model viruses in water hygiene: ecological aspects. Water Sci Tech, 20:399-407.

Havelaar AH, M van Olphen and YC Drost. 1993. F-specific RNA bacteriophages are adequate model organisms for enteric viruses in fresh water. Appl Environ Microbiol, 59(9):2956-2962.

Hsu FC, YS Carol Shieh, J van Duin, MJ Beekwilder and MD Sobsey. 1995. Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl Environ Microbiol, 61(11):3960-3966.

Jiang S, R Noble and W Chu. 2001. Human adenoviruses and coliphages in urban runoff-impacted coastal waters of southern California. Appl Environ Micro 76(1):179-184.

Kirs M and DC Smith. 2007. Multiplex quantitative real-time reverse transcriptase PCR for F+-specific RNA coliphages: a method for use in microbial source tracking. Appl Environ Microbiol, 73(3):808-814.

Lee JV, SR Dawson, S Ward, SB Surman and KR Neal. 1997. Bacteriophages are a better indicator of illness rates than bacteria amongst users of a white water course fed by a lowland river. Wat Sci Tech 35(11-12):165-170.

Lee DY, K Shannon and LA Beaudette. 2006. Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. J Microbiol Meth, 65:453-467.

Leclerc H, S Edberg, V Pierzo and JM Delattre. 2000. Bacteriophages as indicators of enteric viruses and public health risk in groundwaters. A review. J Appl Microbiol, 88:5-21.

Loeb T, and ND Zinder. 1961. A bacteriophage containing RNA. Proc Nat Acad Sci, 47:282-289.

Long SC, SS El-Khoury, SJG Oudejans, MD Sobsey and J Vinjé. 2005. Assessment of sources and diversity of male-specific coliphages for source tracking. Environ Eng Sci, 22(3):367-377.

Long SC, and MD Sobsey. 2004. A comparison of the survival of F+ RNA and F+ DNA coliphages in lake water microcosms. J Water Health, 2(1):15-22.

Love DC and MD Sobsey. 2007. Simple and rapid F+ coliphage culture, latex agglutination, and typing assay to detect and source track fecal contamination. Appl Environ Microbiol, 73(13):4110-4118.

Maynard C, F Berthiaume, K Lemarchand, J Harel, P Payment, P Bayardelle, L Masson and R Brousseau. 2005. Waterborne pathogen detection by use of oligonucleotide-based microarrays. Appl Environ Microbiol, 71(12):8548-8557.

Meschke JS and MD Sobsey. 2003. Comparative reduction of Norwalk virus, poliovirus type 1, F+RNA coliphage MS2 and *Escherichia coli* in miniature soil columns. Water Sci Tech 47(3):85-90.

Miyake T, K Furuse, T Shiba, T Aoi, T Sakurai and I Watanabe. 1971. Isolation and grouping of RNA phages in Taiwan. J Keio Med Soc, 48:25-34. (Japanese).

Miyake T, K Furuse, T Shiba, T Aoi, T Sakurai and I Watanabe. 1973. Isolation and grouping of RNA phages. II. A survey in Brazil. J Keio Med Soc, 50:353-362. (Japanese).

Miyake T, I Haruna, T Shiba, YH Itoh, K Yamane, and I Watanabe. 1971. Grouping of RNA phages based on the template specificity of their RNA replicases. Proc Nat Acad Sci, 68(9):2022-2024.

Miyake T, T Shiba, T Sakurai, and I Watanabe. 1969. Isolation and properties of two new RNA phages SP and FI. Japan J Micro, 13(4), 375-382.

Nelson D. 2004. Phage taxonomy: we agree to disagree. J Bacteriol, 186(21):7129-7031.

O'Connell KP, JR Bucher, PE Anderson, CJ Cao, AS Khan, MV Gostomski and JJ Valdes. 2006. Real-time fluorogenic reverse transcription-PCR assays for detection of bacteriophage MS2. Appl Environ Microbiol, 72(1):478-483.

Ogorzaly L and C Gantzer. 2006. Development of real-time RT-PCR methods for specific detection of F-specific RNA bacteriophage genogroups: application to urban raw wastewater. J Virol Meth, 138:131-139.

Osawa S, K Furuse, MS Choi, A Ando, T Sakurai, and I Watanabe. 1981. Distribution of ribonucleic acid coliphages in Korea. Appl Environ Microbiol, 41(4):909-911.

Osawa S, K Furuse, and I Watanabe. 1981. Distribution of ribonucleic acid coliphages in animals. Appl Environ Microbiol, 41(1):164-168.

Paranchych, W. 1975. Attachment, ejection and penetration stages of the RNA phage infectious process. In: ND Zinder (ed). *RNA Phages*. Cold Spring Harbor Laboratory. pp. 85-111.

Puig A, N Queralt, J Jofre and R Araujo. 1999. Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes from fecally polluted wastewater. Appl Environ Microbiol, 65(4):1772-1776.

Rohwer F and R Edwards. 2002. The phage proteomic tree: a genome-based taxonomy for phage. J Bacteriol, 184(16):4529-4535.

Rose JB, X Zhou, DW Griffin and JH Paul. 1997. Comparison of PCR and plaque assay for detection and enumeration of coliphage in polluted marine waters. Appl Environ Microbiol, 63(11):4564-4566.

Rose JB, SR Farrah, VJ Harwood, AD Levine, J Lukasik, P Menendez and TR Scott. 2004.

Reduction of pathogens, indicator bacteria, and alternate indicators by wastewater treatment and reclamation processes. Water Environment Research Foundation (WERF). Water for Reuse; Final Report. Co-published by IWA.

Sakurai T, T Miyake, T Shiba and I Watanabe. 1988. Isolation of a possible fourth group of RNA phage. Japan J Microbiol, 12(4):544-546.

Schaper M, J Jofre, M Uys, and WOK Grabow. 2002. Distribution of genotypes of F-specific RNA bacteriophages in human and non-human sources of faecal pollution in South Africa and Spain. J Appl Microbiol, 92:657-667.

Sobsey. Coliphage tracking to identify sources of fecal contamination. ftp://ftp.sccwrp.org/

Stewart JR, J Vinjé, SJG Oudejans, GI Scott and MD Sobsey. 2006. Sequence variation among group III F-specific RNA coliphages from water samples and swine lagoons. Appl Environ Microbiol, 72(2):1226-1230.

Stewart-Pullaro, J, JW Daugomah, DE Chestnut, DA Graves, MD Sobsey and GI Scott. 2006. F+RNA coliphage typing for microbial source tracking in surface waters. J Appl Microbiol, 101:1015-1026.

Stoeckel DM. 2005. Selection and application of microbial source tracking tools for water-quality investigations. In: *Techniques and Methods Book 2*, Chapter 3A.

Sundram A, N Jumanial, and MM Ehlers. 2006. Genotyping of F-RNA coliphages isolated from wastewater and river water samples. Water SA, 32(1):65-70.

Tartera C and J Jofre. 1987. Bacteriophage active against *Bacteroides fragilis* in sewage-polluted waters. Appl Environ Microbiol, 53:1632-1637.

Tartera C, F Lucena and J Jofre. 1989. Human origin of *Bacteroides fragilis* bacteriophage present in the environment. Appl Environ Microbiol, 55:2696-2701.

van Duin, J., 1998. The single-stranded RNA bacteriophages. in: Fraenkel Conrat, H., Wagner, R.R. (Eds). The Bacteriophages. Series *The Viruses*. Plenum Press, NY. pp. 117-167.

Vinjé J, SJG Oudejans, JR Stewart, MD Sobsey and SC Long. 2004. Molecular detection and genotyping of male-specific coliphages by reverse transcription-PCR and reverse line blot hybridization. Appl Environ Microbiol, 70(10):5996-6004.

Wade TJ, RL Calderon, E Sams, M Beach, KP Brenner, AH Williams and AP Dufour. 2006. Rapidly measured indicators of recreational water quality are predictive of

swimming-associated gastrointestinal illness. Environ Health Persp, 114(1):21-28.
Wade TJ, N Pai, JNS Eisenberg and JM Colford Jr. 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gstrointestinal illness? A systematic review and meta-analysis. Environ Health Persp 111(8):1102-1109.

Yee, SYF, NY Fong, GT Fong, OJ Tak, GT Hui and YS Ming. 2006. Male-specific RNA coliphages detected by plaque assay and RT-PCR in tropical river waters and animal fecal matter. Int J Environ Health Res, 16(1):59-68.

Zinder ND, RC Valentine, M Roger, and W Stoeckenius. 1963. F1, a rod-shaped male-specific bacteriophage that contains DNA. Virology, 20:638-640.

Zinder ND. 1965. RNA phages. Annu Rev Microbiol, 19:455-473.

V. Gene Mapping and Phylogenetic Analysis of the Complete Genome of 30 ssRNA Male-Specific Coliphages of the Family *Leviviridae*

Abstract

An international collection of male-specific ssRNA (FRNA) coliphages comprising the *Leviviridae* family exists but the genetic diversity of these strains is poorly characterized. FRNA coliphages belonging to the family *Leviviridae* are genetically divided into two genera (*Levivirus* and *Allolevivirus*) which can be further divided into four different genogroups (I, II, III and IV).

To better characterize this virus family and its genetic genogroups, strains were collected from the United States, Japan, Europe, Brazil, Taiwan and/or Mexico. The complete genomic sequences of 19 new FRNA strains (10 *Levivirus* and 9 *Allolevivirus*) from diverse sources were determined and compared to the eleven known genome sequences available in GenBank, for a total of thirty FRNA genomes. Phylogenetic analyses demonstrated all strains clustered into two genera, *Levivirus* and *Allolevivirus*, and in four distinct genogroups, I, II, III and IV. Out of ten genogroup I strains, nine strains shared nucleotide sequence similarities ranging from 91.68% - 99% whereas genogroup I strain fr shared a 75.27-77.65% sequence identity with the other genogroup I phages. Genogroup II FRNA strains shared 83.30-93.84% nucleotide similarity. *Allolevivirus* strains shared a nucleotide similarity range of 69.77-95.69% and 74.90-95.03% for genogroups III and IV, respectively. An approximate 50% nucleotide sequence identity was shared between *Levivirus* groups I and II and between *Allolevivirus* groups III and IV. For all strains

genomic full-length nucleotide and individual protein phylogenetic trees were compared. Genogroup II lysis protein tree formed a unique branch that was not observed in the full-length nucleotide tree. Thus, both full-length nucleotide and individual proteins need to be evaluated when genotyping or phylogenetically clustering these FRNA coliphages. Data for amino acid composition, nucleotide similarities and replicase catalytic domain location contributed to phylogenetic branches or strain subclusters. Eight nucleotides at the 3' termini clearly distinguished between the *Levivirus*, 5' ACCACCCA 3' from the *Allolevivirus* 5' TCCTCCCA 3' genera. This evidence suggests that the sequence data is valid.

Introduction

Male-specific RNA (FRNA) coliphages are single-strand RNA (ssRNA) viruses possessing a positive sense genome ranging from 3.8 to 4.2 kb in size enclosed by a non-enveloped 26 nm icosahedral-shaped capsid (Buchen-Osmond, 2003). The bacterial host is restricted to the gram-negative bacteria of the genera *Escherichia*, *Pseudomonas*, *Caulobacter*, *Salmonella*, and *Vibrio* (Loeb and Zinder, 1961). For successful infection, the host must possess a fertility (F) sex-pilus, coded on the F-plasmid of *E. coli* (Paranchych, 1975), as infection occurs by attachment to this receptor site (Crawford and Gesteland, 1964). As the sex-pili are only expressed at temperatures greater than 30 ^{o} C (Grabow, 2001), phage replication is restricted to environments such as the intestinal tract of warm-blooded animals. The inability to replicate outside the gut, environmental stability, high numbers in sewage and a strong correlation with water-borne pathogens are some of the properties which have

made FRNA phages attractive candidates as indicators of fecal contamination in water (Havelaar et al., 1993).

FRNA phages belong to the family *Leviviridae* and can be further subdivided into two genera (*Levivirus* and *Allolevivirus).* *Levivirus* are subdivided into genogroups I and II and *Allolevivirus* are subdivided into genogroups III and IV. Historically, these subgroups were based on serological properties (Sundram et al., 2006), sedimentation, density and molecular weight (van Duin, 1988). Recently, genomic data has provided an additional subgrouping tool (Stewart et al., 2006).

Based on a limited number of complete sequences four genes could be identified (reviewed by Bollback & Huelsenbeck, 2001). These genes code for an assembly or maturation protein, capsid protein, lysis protein and replicase protein in the *Leviviruses* whereas in *Alloleviviruses,* the lysis protein is replaced by a read-through protein. Each levivirus virion contains one molecule of positive sense ssRNA, 180 copies of the capsid or coat protein, one copy of the maturation protein and, in alloleviviruses, approximately 15 copies of the read-through protein (Weber & Konigsberg, 1975; van Duin, 2000; van Duin and Tsareva, 2006).

A few complete nucleotide sequences of *Leviviridae* strains are known, including prototype strains MS2, GA, Qβ and FI, as only 11 FRNA phages have been fully sequenced over the past few decades. With rapid molecular advances, sequencing is now more affordable and feasible.

Purpose

- Generate a nucleotide (nt) sequence database of complete genomic sequences of representative strains and environmentally isolated strains for all four genogroups of FRNA coliphages.
-
- Determine phylogenetic profiles, nucleotide sequence similarity, amino acid composition, Open Reading Frame (ORF) positions and subsequent gene locations for a total of 30 FRNA sequences.
-

Approach

-
- Sequence nineteen FRNA strains and compare to eleven full-length genome sequences available in GenBank (GenBank/EMBL/DDBJ).
-
- Determine Open Reading Frames by locating Shine-Dalgarno regions and start codons to map each gene.
-
- Translate nucleotide sequences into amino acid compositions for each protein.
-
- Compare nucleotide and amino acid percent similarities among each genogroup.
-
- Determine and compare each protein family, protein motifs and domains by use of bioinformatics tools.
-
- Compare phylogenetic clustering for each full-length nucleotide genome and individual proteins in all genogroups.

Materials and Methods

FRNA Coliphage Strains and RNA Extraction

FRNA strains used in this study include prototype strains MS2 (genogroup I), GA (genogroup II), Qβ (genogroup III), FI (genogroup IV) and SP (genogroup IV) were kindly provided by Dr. K. Furuse (Toaki University, Japan) and FRNA strains ST4, TW18, VK and BZ1 were a gift from Dr. J. van Duin (Leiden University, The Netherlands). Field-collected strains BR1, BR8 and BR12 were a gift from Brian Robinson (NOAA, Charleston, SC) and prototype strain fr was provided by Dr. A. Boehm (Stanford University, Stanford, CA). Strain R17 was purchased from Felix D'Herelle Reference Centre for Bacterial Viruses, Universite Laval, Quebec, Canada. In addition, field strains isolated from wastewater, surface waters, swine lagoons and chicken litter were analyzed in this study (Table 7.1). Preliminary subgrouping of phages was previously determined by reverse line-blot hybridization (Vinjé et al., 2004).

Each strain was plaque purified and further enriched using *Escherichia coli* HS(pFamp)R as host (Vinjé et al., 2004). Single plaques were enriched overnight in Tryptic Soy Broth (TSB) supplemented with streptomycin-sulfate (15 mg/L) and ampicillin (15 mg/L). Cultures were centrifuged (3,220 x g for 10 min) to pellet host cells and debris and the supernatant was chloroform extracted (1:1 V:V). Approximately 1-2 ml aliquots of the purified supernatant were frozen at −75 ^{o} C.

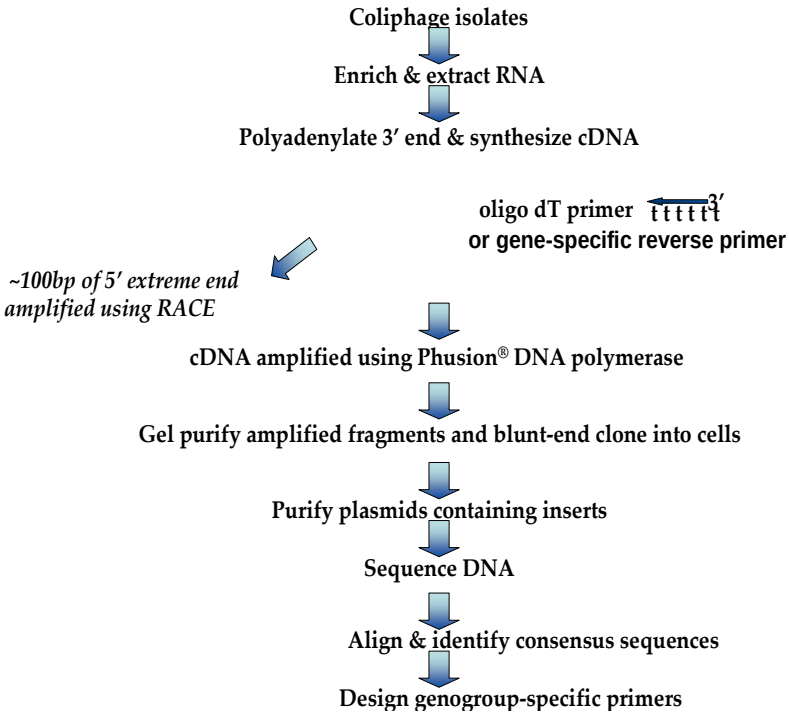
Coliphage titers were determined using a single agar layer procedure (SAL) (US EPA Method 1602, 2001). The procedure was as follows. A 1 ml of overnight *E. coli* Famp host culture was transferred into 50 ml trypticase soy broth with streptomycin and ampicillin

(TSB/strep/amp) and grown 4 hr to log phase. A 150 ml volume of trypticase soy agar, TSA (4.5 g TSB and 1.2 g agar), was autoclaved and placed into a water bath (47-55 ^{o} C). A 300 ul aliquot of 500X strep/amp was added into the cooled 150 ml TSA. Serial 10-fold dilutions of the purified virus were prepared. For 10-fold stock dilution, 100 ul of stock was transferred into 900 ul phosphate buffered saline (PBS), mixed and this process repeated to a serial dilution 10^{-15}. To sterile15 ml plastic tubes, labeled -1 to -15, 1 ml of the 4 hr *E. coli*, 100 ul of the appropriate virus dilution and 10 ml TSA were added. After quick manual mixing the tube contents were poured into labeled 20mm petri plates which were then cooled, allowed to dry, inverted and incubated overnight at 37 ^{o} C. The coliphage titer was determined by counting the plates having a minimum of 5 to a maximum of 20 plaques per plate. Titer concentration was reported as PFU/ml.

Coliphage RNA was extracted from the purified viral supernatant using a QIAamp viral RNA mini kit (Qiagen, Valencia, CA) as follows. A 200 ul volume of purified virus stock was added into the 25 ul protease tube. Into the same tube, 200 ul of carrier RNA/Buffer AL mixture was added and the combined solutions were vortex mixed, incubated 56 ^{o} C for 15 minute and clarified by a 1 minute centrifugation. To the same tube, 250 ul ethanol (EtOH) was added, vortex mixed and incubated 5 min at room temperature (RT). The lysate was applied into the QIAamp spin column and centrifuged 3-4 min at 6000 xg (8000 rpm) and the contents of the collection tube was discarded. To the column, 500 ul of Buffer AW2 was added, centrifuged 1 min and the contents of the collection tube were discarded. To the column, 500 ul EtOH was added, centrifuged 1 min and collection tube was discarded. The column was transferred to a clean collection tube and centrifuged for 1

minute. The column was again transferred a new, nuclease-free tube and 50 ul RNase free water was added onto column and incubated 5 min at RT. The final centrifugation, (14,000 rpm) for 1 min and the recovered purified RNA was frozen at −20 ^{o}C .

Figure 5.1. Flow-chart of sequencing methods.



Generating cDNA from Polyadenylated RNA

For all procedures during cDNA synthesis, strain MS2 was used as a positive control. First, viral RNA was 3' polyadenylated with yeast PolymeraseA and 25 mM ATP in a 50 ul reaction volume (USB, Inc, Cleveland, OH). The reaction was prepared with 10 ul 5X Reaction Buffer, 10 ul RNA, 2 ul 25 mM ATP, 0.7 ul 600 U Poly(A)Polymerase and 27.3 ul nuclease-free water. The mixture was incubated at 37^{0} C for 5 min and placed on ice for

enzymatic termination. Polyadenylated RNA was either immediately frozen or used as a template for cDNA.

Second, full-length cDNA was prepared using oligo-dT reverse primer supplied with the reverse transcriptase MonsterScript 1st Strand cDNA Synthesis Kit (EpiCentre, Madison, WI) or with a gene-specific reverse primer. To a 250 ul thin-walled PCR tube, the following reagents were added: 4 ul nuclease-free water, 10 ul polyadenylated RNA or RNA template and 1 ul of 10 uM PolyT primer or 1 ul of 10uM gene-specific primer. The mixture was heated for 1 min at 65 ^{o} C and chilled for 1 min on ice. To the same tube, 1 ul MonsterScript Reverse Transcriptase and 4 ul of 5X cDNA reaction buffer were added. The mixture was placed in a thermocycler with the following cycle regime: 37 ^{o} C for 5 min, 42 ^{o} C for 5 min, 60 ^{o} C for 40 min. The reaction was terminated by incubating at 90 ^{o} C for 5 min and chilled on ice for 1 min. The single-stranded cDNA was either frozen or used for PCR template (Fig 5.1). To verify the generation of full-length cDNA, a partial region of the the 5' end of MS2 was amplified using primers MS25 and MS23 (Lovmar *et al.*, 2003).

To amplify the 1 kb region between the replicase and the 3' end of the genome, strain-specific forward primers were designed based on a 200 nucleotide (nt) region of the replicase gene (Vinjé et al., 2004). To amplify the 5' end of the genome, reverse primers were designed based on the replicase gene sequence of each strain and forward primers were designed based on available full-length sequences (GenBank) of each genogroup. As sequences were generated (Sequetech, Mountain View, CA), reverse primers were designed

to amplify overlapping sections of the genome. The majority of the genome was sequenced by "primer walking."

5' Amplification of cDNA Ends

The nucleotide sequence of the 5' region was determined by rapid amplification of cDNA end (RACE) with the Smart Race cDNA Amplification Kit (Clontech, Mountain View, CA). First-strand cDNA synthesis was prepared on ice in a 250 ul thin-walled PCR tube by combining 3 ul RNA, 1 ul of 10uM gene-specific reverse primer and 1 ul Smart oligo (from kit). The 5 ul reaction volume was briefly centrifuged and the following components were added: 2 ul of 5X First Strand buffer, 1 ul of 20 mM DTT, 1 ul of 10 mM dNTP and 1 ul SuperScript II (Invitrogen, Carlsbad, CA). Following a brief centrifugation, the mixture was incubated for 90 min at 42 ^{o} C. To dilute the first-strand cDNA, 20 ul of Tricine-EDTA buffer was added and heated for 7 min at 72 ^{o} C. The reaction generated double stranded cDNA. The cDNA was frozen at - 20 ^{o} C and used for subsequent PCR reactions. Concentration of cDNA was determined with a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Long Template PCR, Cloning and Sequencing

The cDNA was amplified by using Phusion DNA Polymerase (New England Biolabs, Ipswich, MA), with final concentrations of 1X of 5X Phusion Buffer, 0.2 mM dNTP, 1 ul of 10 uM forward primer, 1 ul of 10uM reverse primer, 3% DMSO, 2 ul cDNA and 0.5 ul Phusion Taq in a 50 ul reaction using the following cycle parameters: one round denaturation at 98 ^{o}C (1 min), 35 rounds at 98 ^{o}C (30 sec), 48 ^{o}C (1 min), 72 ^{o}C (3 min) followed by 10 min extension at 72 ^{o}C. For each reaction, positive controls were prepared using primers

MJV82 and JV81 for *Leviviruses* and MJV82 and JV41 with *Allolevivirus*es (Vinjé et al., 2004). A no-template negative control was included.

PCR products were separated by electrophoresis in a 1.5% agarose gel, stained with SYBR Gold Nucleic Acid Gel Stain (Molecular Probes, Carlsbad, CA) and visualized under blue light (Dark Reader Transilluminators, Clare Chemical Research, Dolores, CO).

Blunt-end PCR products were excised using a gel extraction tool (USA Scientific Plastics, Ocala, FL) and then purified (QuickClean 5M Gel Extraction Kit, GenScript Corporation, Piscataway, NJ). Excised bands were weighed in 1.5 ml microcentrifuge tubes and 3 volumes of Binding Solution II per gel slice were added. The gel solution was heated at 50 ^{o} C until melted. One volume of isopropanol was added and the mixture was transferred into the Genprep spin column and centrifuged for 1 min at 12,000 rpm. The column effluent was discarded and 500 ul of Wash Buffer was added, the column was centrifuged and the liquid waste discarded. The 500 ul wash was repeated and waste discarded. The column was placed into a clean 1.5 ml microcentrifuge tube, 30 ul of Elution Buffer was added and the column was incubated 1 min at room temperature. The tube was centrifuged 1 min at 12,000 rpm and the collected DNA eluate was transferred to a clean tube. Concentration of the DNA was determined using a NanoDrop spectrophotometer. The DNA was either cloned or the PCR product was sequenced.

Gel-purified DNA was cloned using a ZeroBlunt TOPO Cloning Kit, pCR-Blunt II TOPO plasmid kit (Invitrogen, Carlsbad, CA). To prepare a TOPO cloning reaction, 4 ul DNA, 1 ul salt solution and 1 ul pCR II Blunt TOPO were added to a nuclease-free tube and incubated 30 min at room temperature. To a vial of One Shot *E. coli* competent cells, 2 ul of

the TOPO cloning reaction were added and incubated on ice for 30 min. The cells were heat-shocked at 42 ^{o} C and immediately placed on ice. A 250 ul volume of SOC medium was added to the One Shot cells and shaken (200 rpm) for 1 hr at 37 ^{o} C. Fifty ul of transformed cells were plated onto pre-warmed LB agar plates containing 50 ug/ml kanamycin, the transformed cells were spread to isolate colonies and the plates were incubated overnight at 37 ^{o} C.

Colonies of transformed *E. coli* cells were screened for positive inserts using whole-cell PCR. Using aseptic techniques, individual colonies were selected with a sterile toothpick, the toothpick was briefly rinsed into a 50 ul Phusion master mix (as described above) then dropped into 10 ml LB/kan broth and incubated overnight at 37 ^{o} C. Whole-cell PCR was performed on individual colonies using Phusion DNA Polymerase and the same primers used to generate the pre-cloned amplicon with the following cycle modifications: one round denaturation at 98 ^{o} C (3 min), 35 rounds at 98 ^{o} C (10 sec), 57 ^{o}C (30 sec), 72 ^{o} C (30 sec) followed by 10 min extension at 72 ^{o} C. Amplicons were separated by electrophoresis in 1.5% agarose gel in 0.5X Tris-acetate-EDTA (TAE), stained with 20 ug/ml ethidium bromide and visualized under UV light (UVP, Upland, CA). Those clones with the appropriate size PCR amplicon were selected for plasmid purification.

Positive clones were plasmid-purified (QIAprep Spin Miniprep Kit, Qiagen, Valencia, CA). An *E. coli* colony that had been selected with a toothpick and incubated overnight was centrifuged 10 min at 8000 rpm (6800 xg). The supernatant was discarded and the cell pellet was resuspended and processed as follows. A 250 ul volume of Buffer P1 was added to the cell pellet, vortexed to mix and transferred to a clean 1.5 ml

microcentrifuge tube. A 350 ul volume of Buffer P2 was added to the resuspended pellet and mixed by inversion followed by addition of 350 ul Buffer N3. The buffer mixture was inverted 4-6 times and centrifuged 10 min at 13,000 rpm (17,900 xg). The supernatant was decanted into the QIAprep spin column and centrifuged 1 min at 13,000 rpm. The column effluent was discarded and the column was washed with 500 ul Buffer PB, the column was centrifuged 1 min and the wash was discarded. To the spin column, 750 ul Buffer PE was added, the column was centrifuged and the wash was discarded. The column was incubated with 50 ul Buffer EB to elute the plasmid. Following a 1 min incubation at room temperature, the column was centrifuged 1 min. Purified plasmid was shipped frozen for sequencing. Each cDNA PCR amplicon was cloned and sequenced in triplicate to obtain the consensus sequence. In some cases, PCR products were sequenced directly. To achieve publication quality data, both forward and reverse strands were sequenced (Sequetech, Mountain View, CA). This process was repeated until complete genomes were obtained.

To avoid contamination, a PCR hood (AirClean 600, AirClean Systems, Raleigh, NC) located in a designated room was used to prepare master mixes separate from template additions. PCR amplification, electrophoresis, template and/or viral preparations (EPA, 2004) were conducted in individual assigned rooms based on designated use.

Sequence Analyses

Raw sequences from three to five individual clones were imported and aligned using BioEdit v7.0.1 (Hall, 1999) followed by Basic Local Alignment Search Tool (BLAST, National Center for Biotechnology Information) analyses for sequence and phylogenetic confirmation. Completed sequences from all strains were aligned with full-length prototype

strains (GenBank) using BioEdit ClustalW application. Open Reading Frames (ORF) for each strain was determined using BioEdit.

Similarity analyses were evaluated using SimPlot v3.5.1 (Lole et al., 1999). Relationships among aligned nucleotide sequences were depicted in similarity plots. The SimPlot program determines the percent identity between a reference sequence and the queried sequence. Percent similarity was calculated within a sliding window 160 bp wide with a step size of 10 bp between plots.

Amino Acid Analysis

Deduced amino acid sequences for each of the four genes were determined using a computer-generated DNA-to-protein translation tool, ExPASY (http://ca.expasy.org/). Prediction of protein sequence motifs were identified by PROSITE (http://ca.expasy.org/) and protein families and domains were modeled in Pfam (http://pfam.janelia.org).

Genetic distance was calculated for each protein within their respective genogroup as follows. Protein amino acid composition was aligned using BioEdit ClustalW followed by protein distance matrix using Neighbor Joining analysis (BioEdit). Matrices values are the fraction of mismatches at aligned positions. Protdist (BioEdit) protein distance matrix compares the number of amino acid mismatches within each protein. Therefore, the smaller the distance value, the higher the amino acid similarity.

Phylogenetic Analysis

Sequence data were analyzed using BioNumerics Software v.3.5 (Applied Maths, Saint-Martens-Latem, Belgium). Phylogenetic trees were built by global cluster analysis performed on multiple aligned sequences and clustered by UPGMA using the Jukes and

Cantor correction (Jukes & Cantor, 1969). A bootstrap analysis, based on 10,000 substitutions, was used to measure cluster significance. The reliability of each cluster was expressed on a percentage basis.

Nucleotide Sequence Accession Numbers

The accession numbers of full-length leviviruses sequences available in GenBank were as follows: Genogroup I MS2 (NC_001417.1), M12 (AF195778), fr (NC_0011333.1); Genogroup II GA (NC_001426.1), KU1 (NC_002250.1); Genogroup III M11 (NC_004304.1), Qbeta (AY099114.1), MX1 (NX_001890.1); and Genogroup IV SP (X07489.1), NL95 (AF059243.1), FI (EF068134.1).

Results

Comparison of Full-length Genome Sequences

Full-length genome sequences of 19 FRNA strains were determined in this study and compared to 11 strains previously published in GenBank with respect to genome size and Open Reading Frame(s) locations (Table 5.1).

Nucleotide sequence similarities among the *Leviviridae* strains are shown in Table 5.2. A total of 7 group I strains were sequenced, DL1, DL2, DL13, DL16, ST4, R17 and J20 and compared to GenBank group I strains MS2, M12 and fr. Group I strains DL2 and DL13 were omitted from this table as they were >99 % identical having only 4 nt single-point mismatches to DL16. MS2 and ST4 were 98.71% similar to each other. Sequence similarity among genogroup I strains ranged from 75.27 - 96.67 % with strain fr forming a separate subgroup (Table 5.2, Fig 5.2).

Table 5.1 Open Reading Frames positions for *Leviviridae*.

		Open Reading Frame Location (nt)				
Strain	Group	Full length nt	ORF1	ORF2	ORF3	ORF4
DL1	I	3570	130-1311	1335-1727	1678-1905	1761-3398
DL2	I	3491[b]	130-1311	1335-1727	1678-1905	1761-3398
DL13	I	3491[b]	130-1311	1335-1727	1678-1905	1761-3398
DL16	I	3569	130-1311	1335-1727	1678-1905	1761-3398
J20	I	3569	130-1311	1335-1727	1678-1905	1761-3398
ST4	I	3569	130-1311	1335-1727	1678-1905	1761-3398
R17	I	3569	130-1311	1335-1727	1678-1905	1761-3398
MS2[a]	I	3569	130-1311	1335-1727	1678-1905	1761-3398
M12[a]	I	3340[b]	130-1311	1335-1727	1678-1905	1761-3340[b]
fr[a]	I	3575	129-1310	1336-1728	1691-1906	1762-3399
GA[a]	II	3465	136-1308	1325-1717	1717-1908	1749-3347
KU1[a]	II	3486	137-1309	1325-1717	1683-1929	1770-3368
T72	II	3393[b]	137-1309	1325-1717	1683-1916	1770-3368
DL10	II	3376[b]	136-1309	1326-1718	1715-1906	1747-3345
DL20	II	3458	137-1309	1326-1718	1718-1909	1750-3348
TW18	III	4218	62-1324	1345-1746	1345-2334	2344-4122
HL4-9	III	4221	62-1324	1345-1746	1345-2337	2347-4125
BR12	III	4218	62-1324	1345-1746	1345-2334	2344-4122
VK	III	4218	62-1324	1345-1746	1345-2334	2344-4122
BZI	III	4219	62-1324	1346-1747	1346-2335	2345-4123
QB[a]	III	4215[b]	56-1319	1339-1741	1339-2329	2338-4117
M11[a]	III	4217	57-1322	1344-1744	1344-2333	2352-4118
MX1[a]	III	4215	56-1321	1343-1744	1343-2332	2351-4111
HP-P22	IV	4241	52-1374	1395-1793	1395-2387	2407-4137
HP-P24	IV	4243	53-1378	1397-1795	1397-2389	2409-4139
BR1	IV	4273	52-1404	1424-1822	1424-2419	2439-4169
BR8	IV	4273	52-1404	1424-1822	1424-2419	2439-4169
NL95[a]	IV	4248	53-1318	1402-1800	1402-2394	2414-4144
SP[a]	IV	4276	55-1407	1427-1825	1427-2422	2442-4172
FI[a]	IV	4184[b]	55-1371	1392-1791	1392-2391	2406-4167

nt = nucleotide; [a] previously published GenBank genomes; [b] nearly full-length genome

Three environmental group II strains, DL10, DL20 and T72 were sequenced and

compared to GenBank group II GA and KU1. Among group II strains nucleotide sequence similarity ranged from 83.30 to 93.84% with strains DL10, DL20 and GA having the highest sequence identities (93.43-93.67%) whereas strains T72 and KU1 formed a separate subcluster (Table 5.2, Fig 5.2). Strains in group I had only 50% sequence identity (range of 46.74-53.85%) with strains in group II (Table 5.2, Fig 5.3A). However, all *Levivirus* strains shared an eight nucleotide sequence at the 3' terminus, 5' ACCACCCA 3'.

Among *Allolevivirus* group III, two different subclusters were formed. The first subcluster was composed of strains VK, HL4-9, BR12, BZ1, TW18 and GenBank strain Qβ having a nucleotide sequence similarity ranging from 91.87-95.69%. The second subcluster formed with GenBank group III strains MX1 and M11 having an 87% nucleotide similarity to each other. The nucleotide similarity of strains between the two group III subclusters ranged from 69.77-71.33%. Group III strains shared < 40% identity (29.73-39.06%) to *Levivirus* groups I and II (Table 5.2). *Allolevivirus* strains share the 3' terminus signature sequences 5' TCCTCCCA 3' (Table 5.5).

Genogroup IV environmental strains BR1, BR8, HB-P22 and HB-P24 were sequenced and compared to GenBank group IV strains SP, FI and NL95. Group IV *Allolevivirus* shared sequence identities ranging from 74.90-95.03 % with the closest identities being 95.03% between strains BR8 and BR1. Strain HB-P24 shared 90.22% identity with prototype strain NL95, whereas strains BR8 and BR1 shared a greater percent similarity with prototype strain SP (91.05-91.73%). Group IV sequence identity was 53.53-57.99% when compared to *Allolevivirus* group III (Table 5.2, Fig 5.3B) and < 40 % (31.81-

38.73%) when compared to *Levivirus* groups I and II.

Table 5.2 Leviviridae nucleotide percent similarity.

Number/ Strain/ Group	% Nucleotide sequence similarity with the indicated strain																										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1/ MS2/ I	100.00																											
2/ ST4/ I	98.71	100.00																										
3/ R17/ I	96.39	96.67	100.00																									
4/ M12/ I	92.78	92.93	92.93	100.00																								
5/ J20/ I	92.19	92.47	92.38	91.82	100.00																							
6/ DLI/ I	92.16	92.41	92.16	92.48	93.95	100.00																						
7/ DL16/ I*	91.68	92.07	91.96	92.00	94.73	94.79	100.00																					
8/ fr/ I	75.27	77.09	75.89	75.43	76.63	76.55	75.65	100.00																				
9/ DL10/ II	53.13	52.18	52.54	52.29	53.47	53.15	52.35	49.93	100.00																			
10/ DL20/ II	52.89	52.23	52.17	52.18	53.85	50.65	53.69	51.68	93.84	100.00																		
11/ T72/ II	52.37	52.35	52.49	50.41	51.42	49.99	49.93	46.74	85.97	85.17	100.00																	
12/ GA/ II	51.35	51.44	51.76	51.75	51.14	50.17	51.79	48.25	93.67	93.43	84.79	100.00																
13/ KU/ II	49.55	49.61	49.69	49.72	51.15	50.23	49.58	48.71	84.69	83.83	89.89	83.30	100.00															
14/ VK/ III	38.66	34.85	37.90	33.79	34.59	37.31	36.13	33.01	37.39	37.37	36.40	37.61	36.39	100.00														
15/ HL4-9/ III	37.55	36.25	37.05	36.19	35.25	32.75	36.68	36.13	34.16	38.11	36.51	34.38	39.25	91.97	100.00													
16/ QB/ III	36.87	36.33	36.94	36.15	38.27	34.07	37.88	32.05	38.02	38.44	36.58	34.85	36.62	92.09	94.76	100.00												
17/ BR1/ III	36.85	36.91	36.91	36.47	37.17	36.32	36.19	32.69	37.95	38.49	36.83	38.73	36.44	95.69	91.87	91.97	100.00											
18/ BZ1/ III	36.79	36.91	37.61	36.31	36.68	37.05	36.53	37.00	31.95	32.27	36.70	33.04	39.06	93.37	92.30	92.24	92.87	100.00										
19/ TW18/ III	36.39	36.73	36.39	35.97	35.28	32.71	36.19	31.71	37.63	38.22	36.13	34.73	36.31	92.25	95.90	95.58	92.06	92.65	100.00									
20/ SP/ IV	36.29	34.64	34.73	34.64	34.90	34.77	34.18	33.77	37.09	36.69	32.80	37.13	32.32	55.49	57.75	54.19	54.64	53.09	55.92	100.00								
21/ HB-P24/ IV	35.39	35.47	34.70	35.43	37.00	33.64	38.54	35.85	37.30	33.03	37.54	36.81	38.73	56.13	55.93	53.53	56.73	54.87	55.89	79.37	100.00							
22/ BR1/ IV	34.13	34.13	36.77	34.61	34.59	37.12	34.30	37.55	37.21	33.08	32.77	33.18	31.81	53.61	57.37	53.93	55.36	54.31	56.94	91.73	79.60	100.00						
23/ FI/ IV	34.05	33.96	32.90	37.03	33.81	37.93	37.63	33.20	31.99	33.41	33.07	33.30	32.53	54.43	56.67	55.65	54.85	55.98	56.15	76.63	76.85	74.90	100.00					
24/ BR8/ IV	34.04	36.61	34.19	33.43	34.70	36.68	34.76	33.66	36.66	36.19	36.83	36.91	36.92	55.27	56.64	53.17	55.33	55.49	54.82	91.05	78.87	95.03	74.57	100.00				
25/ M11/ III	33.47	31.11	33.49	30.54	35.66	32.79	32.95	36.14	33.59	33.47	31.82	33.75	30.57	71.33	70.43	70.45	71.05	70.66	71.23	53.67	55.84	52.33	56.03	51.61	100.00			
26/ NL95/ IV	33.10	37.11	32.64	32.55	36.94	32.75	32.87	33.46	36.05	36.04	37.76	37.58	36.99	55.99	56.39	55.81	55.80	54.64	53.92	78.50	90.22	78.47	76.07	78.32	55.89	100.00		
27/ MX1/ III	32.53	32.33	31.21	29.73	36.44	37.49	32.27	33.09	33.47	32.59	37.81	33.45	32.81	70.27	70.55	69.90	70.13	69.77	70.92	52.00	55.08	52.46	54.03	51.19	87.15	54.77	100.00	
28/ HB-P22/ IV	32.30	32.87	32.39	32.28	33.43	32.55	32.65	33.06	36.29	37.01	37.56	36.69	37.54	57.79	57.99	57.37	57.65	56.32	57.34	81.80	82.83	82.07	77.69	81.38	54.97	82.25	54.43	100.00

*DL2 and DL13 were omitted. DL16 was used as a representative strain as all three strains differed by only 4 nucleotides.

Figure 5.2 Phylogenetic analysis of *Leviviridae* nucleotide sequences.

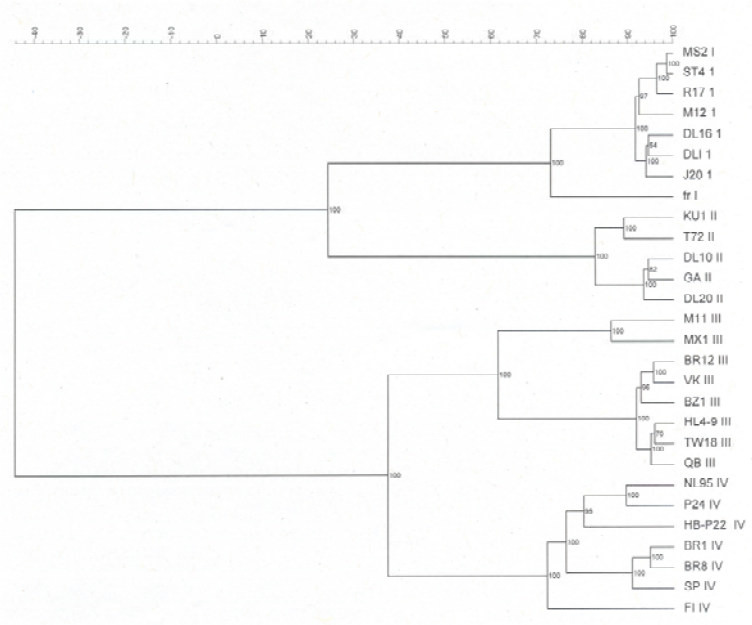
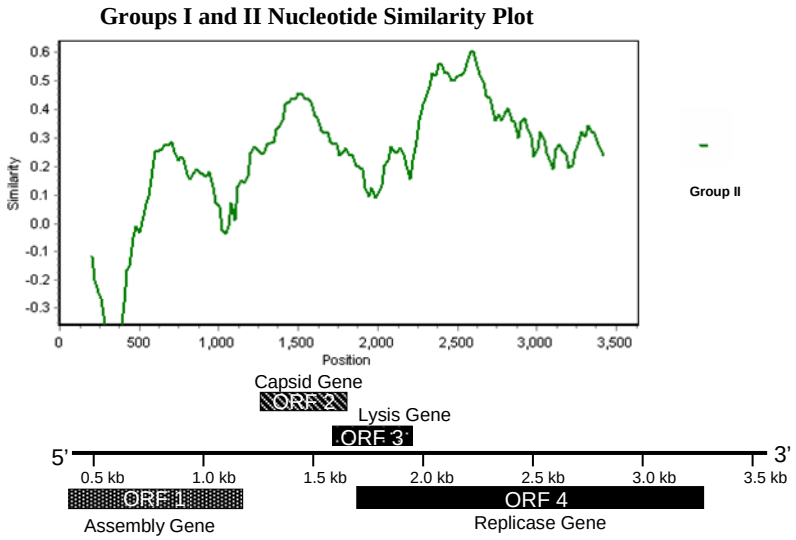


Figure 5.3 SimPlot nucleotide similarity and genome organization.



Open Reading Frame Analyses

Genome lengths, in nucleotides (nt), gene and Open Reading Frame (ORF) locations of all sequenced FRNA phages belonging to genus *Levivirus* and genus *Allolevivirus* are shown in Table 5.1. Shine-Dalgarno sequences and start codons for all genes of these phages are presented in Table 5.3.

Levivirus

For all group I strains except strain fr, the ORF start and stop codons were located at the same nucleotide (nt) positions as was previously reported for strain MS2 (Fiers et al., 1976). Four proteins and four ORFs were located at the following nt positions: maturation/assembly protein (130-1311), coat protein (1335-1727), lysis protein (1678-1905) and replicase (1761-3398) (Table 5.1). In contrast, the ORF positions for strain fr were as follows: maturation protein A (129-1310), coat protein (1336-1728), lysis protein (1691-1906) and replicase (1762-3398). The start codon for ORF1, ORF2, ORF3 and ORF4 of genogroup I were AUG with two exceptions (1) MS2, ST4 and fr start codon for ORF1 was GUG and (2) lysis gene start codon, ORF3, for strain fr was UUG (Table 5.3).

Group II lysis and replicase genes were more similar in nucleotide position among strains T72 and KU1 whereas strains DL10 and DL20 were similar to strain GA (Table 5.1, Table 5.3). Group II start codons for all ORFs were AUG (Table 5.3). The four genes in group II differed from group I in their ORF nucleotide positions. Lengths of the 5' untranslated regions of groups I and II genomes were 128 to 129 nt and 135-136 nt length, respectively. Lengths of the 3' untranslated regions were 168-177 nt in group I and 135-140 nt length in group II (Appendix A).

Table 5.3 Start condons and Shine-Dalgarno sequences. Start codons (bold, underlined) and Shine-Dalgarno (underlined) sequences for each gene. (A) alignment of strains DL1, DL2, DL13, DL16, J20, ST4, R17 and strains M12 and MS2 from Group 1 Levivirus. Strain fr is shown separately. (B) alignment of strains T72, DL10, DL20, and strains KU1 and GA from Group II Levivirus, (C) alignment of strains TW18, HL4-9, BR12, BZ1, VK and strains QB, MX1 and M11 from Group III Allolevivirus, (D) alignment of strains HB-P22, HB-P24, BR1, BR8 and strains SP, NL95 and FI from Group IV Allolevivirus.

(A). Levivirus Group I.

ORF	Gene	Sequences
1	assembly	CCAUUCCUAGGAGGUUUGACYY**RUG**CGAGC
	strain fr	GCUAGGGAGCCUCGU**GUG**CGAAAGU
2	capsid	AACCGGAGUUYGAAGC**AUG**GCUUCUAA
	strain fr	CCGAAGGGAGAGCCAC**AUG**GCUUCG
3	lysis	RAUGCAAGGUCUCCURAAAG**AUG**GAAACCC
	strain fr	AACUGGUAACCCAA**UUG**CAACAGC
4	replicase	CAUGAGGAUUACCC**AUG**UCGA
	strain fr	ACAUGAGGAAUACCC**AUG**UCAAAAU

(B) Levivirus Group II.

ORF	Gene	Sequences
1	assembly	AUACCGGAGGADCU**AUG**UUUCCGA
2	capsid	AAWWAYGGAGUUAGCCAY**AUG**GCAACUUUA
3	lysis	
	strains T72, KU1	GAUUGGGAACCCGGUUGCUG**AUG**CCAUCUC
	strains DL10, DL20, GA	CDCAGAGYGGCUUCUACGCGUA**AUG**GGUCUG
4	replicase	
	strains T72, KU1	CAUAAGGAAAACCU**AUG**UUCCGAUUCA
	strains DL10, DL20, GA	AAACAUAAGGAAAACCU**AUG**UUCCGAUUCA

(C) Allolevivirus Group III.

ORF	Gene	Sequences
1	assembly	DRGAGGMMAY**AUG**CCWM
2/3	capsid/readthrough	UUGGGUCAAUUHGAUC**AUG**GCWAAA
4	replicase	
	TW18, HL4-9, BR12, BZ1, VK, QB	AGUAACUAAGGAUGAAAUGC**AUG**UCUAAG
	MX1 and M11	AGUAACURAAGGAGAUCUGC**AUG**UCWA

(D) Allolevivirus Group IV.

ORF	Gene	Sequences
1	assembly	CUACAGAGGAGAAUCU**AUG**CC
2/3	capsid/readthrough	
	strains BR1, BR8, NL95, SP	CUUUGGGUCAAUUYGAUC**AUG**GCAA
	strains HB-P22, HB-P24, FI	YUUUGGGUCAAUUYGAUC**AUG**GCWA
4	replicase	
	strains BR1, BR8, NL95, SP	CUUAARRGAGRWAGC**AUG**YCAA

Allolevivirus

The *Allolevivirus* genome possesses four genes and 3 start codons as the capsid and readthrough genes share a single ORF (ORF2/ORF3).

For all group III strains, except MX1 and M11, the ORF alignment positions were very similar or identical. Although ORFs of strain Qβ aligned perfectly with the other group III strains, the GenBank acquired Qβ sequences were not complete. Thus, individually mapped ORF positions varied slightly (Table 5.1). MX1 and M11 ORF and Shine-Dalgarno positions were similar to the other group III strains for the assembly, capsid and readthrough genes, but differed for the replicase gene (Table 5.3).

Within group IV phages BR1, BR8, NL95 and SP had similar ORF positions whereas strains HB-P22, HB-P24 and FI ORFs had similar nt positions (Table 5.1, Table 5.3).

The 5' untranslated length in groups III and IV ranged from 56-62 nt and 50-53 nt, respectively. Groups III and IV 3' untranslated sequence length were 96-98 nt and 97-104 nt, respectively.

Shine-Dalgarno Sequences

Shine-Dalgarno sequences for *Levivirus* groups I and II were located within 5-9 nt upstream from ORF1 and ORF2, 9-16 nt upstream from ORF3 and 7-8 nt upstream from ORF4. Shine-Dalgarno sequences for *Allolevivirus* groups III and IV were located 5-6 nt upstream from ORF1, 12 nt upstream from ORF2/ORF3, and 6-9 nt upstream from ORF4 (Table 5.3).

With respect to the entire *Leviviridae* family, the ORF positions for the coliphage genes were preceded by Shine-Dalgarno sequences located within 5-16 nt upstream from the ORF start codon(s).

Amino Acid Composition - *Levivirus*

The numerical amino acid (aa) positions are denoted as to their location within their respective protein. *Levivirus* group I demonstrated the most conservative amino acid composition when compared to groups II, III and IV (Table 5.4, Appendix B).

With the exception of strain fr, amino acid number was consistent in each of the four protein types in group I phages (Table 5.4, Appendix B). In the lysis protein, strain fr had 71 amino acids, a four codon deletion, when compared to 75 amino acids in the other group I strains. Alignment of the group I lysis protein, including strain fr, indicated 2 stretches of conserved amino acids; one region of 11 amino acids and another region having 13 amino acids. A 15 aa deletion was observed in group II lysis protein strains DL10, DL20, TL2 and GA when compared to T72 and KU1. No consensus sequence was noted in the lysis protein between groups I and II. The capsid protein of all strains in group I was 130 amino acids in length. *Levivirus* groups I and II capsid proteins shared a conserved region consisting of a 10 amino acid, FVLVDNGGTG, consensus sequence. Groups I and II maturation protein shared a consensus region RWLELQ at amino acid positions number 198-203. Groups I and II replicase shared a 24 amino acid conserved region at positions 198-221 and the YDGG sequence near amino acid position 340 (Appendix B).

Table 5.4 *Leviviridae* number of amino acids per gene. Male-specific coliphage family *Leviviridae*, genera *Levivirus* (groups I and II) and *Allolevivirus* (groups III and IV) table of amino acids. Proteins were determined by mapping of Open Reading Frames and translation of nucleotide sequences to amino acids using ExPASY (http://ca.expasy.org/) DNA-to-protein translation tool. The number of amino acids per protein is listed for each genogroup. If differences existed within groups, the amino acid numbers are listed for individual strains. (A) Group I, (B), Group II, (C) Group III and (D) Group IV.

	Maturation	Capsid	Lysis	Replicase
(A) Group I strains				
MS2, M12, DL1, DL2, DL13, DL16, J20, ST4, R17:	393	130	75	545
fr:	393	130	71	545
(B) Group II strains				
GA, DL10, DL20:	390	130	63	532
T72:	390	130	75	532
KU1:	390	130	79	532
	Maturation	Capsid	Read-thru	Replicase
(C) Group III strains				
QB, BR12, TW18, BZ1, VK:	420	133	328	592
HL4-9:	420	133	329	592
MX1:	421	133	328	586
M11:	421	133	328	588
(D) Group IV strains				
SP, BR1, BR8:	450	132	330	576
FI:	438	132	332	586
NL95:	442	132	331	576
HB-P24:	441	132	329	576
HB-P22:	440	132	329	576

Pfam - *Levivirus*

Not all of the four genes, as determined by ORF mapping, were grouped into both a protein domain and family membership by Pfam. A PfamA maturation protein search generated "phage_mat-A" domain and a total of 24 *Leviviridae* phages including all four genogroups and the additional species PRR1, PP7 and AP205. *Levivirus* capsid amino acid compositions were placed into the "Levi_coat" family and the replicase protein was placed into the "RNA replicase, beta-chain" domain. The group I lysis protein was not sorted to a family or domain in a PfamA search. A subsequent PfamB search for the group I lysis protein linked it to a lysis domain and the results generated *Levivirus* group I species fr, M12, MS2 and JP501.

For each protein, Pfam recognized *Leviviridae* ssRNA viral species. In some cases, such as the group I capsid protein, GenBank *Leviviridae* strains from groups I, II, III, IV and bacteriophage PRR1 were included in the Pfam species tree. In addition to the *Leviviridae* family, results of searches for the groups I and II replicase species tree added bacteriophages PRR1, ZR, BO1 and Acinetobacter phage AP205.

Protein sequence motifs - *Levivirus*

Predicted protein motifs, casein kinase II phosphorylation, cAMP and cGMP-dependent protein kinase phosphorylation, protein kinase C phosphorylation, N-myristoylation, N-glycosylation and tyrosine kinase phosphorylation, occurred frequently in the FRNA coliphages. Excluding strain fr, the maturation protein of groups I and II shared one amino acid motif with the casein kinase II phosphorylation protein and the replicase gene shared one amino acid position motif with the N-myristoylation protein.

No amino acid motif positions were shared between groups I and II in the capsid or lysis proteins. Unique to strain fr was the presence of a leucine zipper in the lysis protein and an amidation motif in the replicase region. The replicase gene RNA-dependent RNA polymerase catalytic domain occurred at amino acid positions 243-373 and 245-375 for groups I and II, respectively. Common to every group II strain was a prenyl group binding site (CAAX box) at amino acid positions 529-532 in the replicase region.

Genetic distances - *Levivirus*

Excluding strain fr, group I amino acid compositions were very similar as the genetic distance was very close among all four proteins. The capsid protein was most similar (distance of 0.0000 - 0.0411) followed by replicase (0.0033-0.0887), maturation protein (0.0046-0.0889) and lysis (0.0232-0.3416). Capsid protein was identical among strains DL1, DL2, DL13, DL16 and J20 (distance 0.0000). Amino acid composition of all four strain fr proteins showed the greatest distance from the other group I strains (0.2316-0.5685).

Closest to furthest genetic distance in the amino acid composition for the group II strains were as follows: capsid (0.0135-0.1116), maturation (0.0235-0.2036), replicase (0.0415-0.2160) and lysis (0.0500-4.6735).

Amino Acid Composition - *Allolevivirus*

The length of the maturation protein of groups III and IV varied from 420 aa to 450 aa (Table 5.4) and possessed a mutual conserved aa region LWLEFRYGL (Appendix B). The length of the capsid protein was 133 and 132 aa for groups III and IV, respectively, and conserved stretches of amino acids occurred in both groups. Groups

III and IV read-through protein was 328 to 332 aa in length and possessed conserved regions between aa positions 290-310. Consensus, however, was unique to each group. Compared to the other group III strains MX1 and M11 shared a three amino acid deletion at the 5' end. The group IV maturation protein in strains HB-P22, HB-P24, NL95 and FI had a 9 amino acid deletion when compared to the other group IV strains. The *Allolevivirus* replicase gene was 576-592 amino acids in length revealing the longest region of conserved amino acids ranging from amino acid positions 202-247 in group IV and positions 207-240 in group III. Groups III and IV replicase shared the sequence KAVTVPKNSKTDRCIAIEPGWNMFFQL in the 210-235 aa region and the YGDD sequence near amino acid position 360 (Appendix B).

Pfam - *Allolevivirus*

Identical results were obtained for groups III and IV PfamA and PfamB searches. Similar to the *Levivirus*, a maturation protein search generated "phage_mat-A" domain and a total of 24 *Leviviridae* phages including all four genogroups and the additional bacteriophage strains PRR1, PP7 and AP205. PfamA displayed the capsid protein in the family "Levi_coat" and linked to a 31-member *Leviviridae* species tree which included all four genogroups along with phages ZR, TH1, TL2, SD, f2 and BO1 plus bacteriophage strains PRR1. Strains PP7 and AP205 were not included in the capsid species tree.

Read-through proteins were grouped as "A1-protein coat readthrough" with PfamB generating a 5-member species tree of SP, QB, NL95, MX1 and M11. As with the *Levivirus*, the *Allolevivirus* replicase protein was sorted into the "RNA replicase, beta-

chain" family including a 24-member *Leviviridae* species tree with additional bacteriophages PRR1, ZR, PP7, BO1 and AP205.

Protein sequence motifs - *Allolevivirus*

As observed in the *Levivirus* genus, the most prevalent *Allolevivirus* protein motifs were casein kinase II phosphorylation, cAMP and cGMP-dependent protein kinase phosphorylation, protein kinase C phosphorylation, N-myristoylation, N-glycosylation and tyrosine kinase phosphorylation. With the exception of group III strains MX1 and M11 and group IV strain HB-P24, a cell attachment motif (RGD) in the maturation protein was present in the majority of group III and IV strains. Group IV strains SP, BR8, BR1 and HB-P22 had a cell attachment motif in the read-through protein.

The catalytic domain of the RNA-dependent RNA polymerase (replicase protein) was located at amino acid positions 262-394 in group III strains except for M11 and MX1. Their catalytic domain was located at amino acid positions 259-391. Group IV catalytic domain was located at amino acid positions 259-391.

Genetic distances - *Allolevivirus*

Group III amino acid genetic distances were most similar in the capsid protein (distance 0.0000-0.3734) followed by replicase (0.0278-0.6571), maturation protein (0.0347-0.8289) with the greatest genetic distance in the read-through protein (0.0444-0.5128). Strains BR12 and VK shared identical capsid proteins (distance of 0.0000).

In group IV, highly similar amino acid compositions were found in the replicase

(distance 0.0474-0.3382), followed by the capsid (0.0535-0.2569) and read-through protein (0.0555-0.5072). The greatest genetic distance was observed in the maturation protein (0.0607-0.5646) in group IV strains.

Phylogenetic Analyses

An algorithmic approach was selected to construct phylogenetic trees from the nucleotide sequence and amino acid data (Fig 5.2, Fig 5.4). Nucleotide sequences in the phylogenetic tree of *Levivirus* group I strains produced two branches, with 9 strains clustered as MS2-like and strain fr an individual branch (Fig 5.2). Within group II nucleotide sequences, strains KU1 and T72 formed one branch and strains DL10, DL20 and GA formed a second branch. *Allolevivirus* group III nucleotide sequences clustered into a MX1, M11 branch and a second branch with Qβ-like strains BR12, VK, BZ1, HL4-9, TW18 and prototype Qβ. Nucleotide sequence analysis formed three branches in group IV strains as follows: 1) HB-P24, HB-P22 and prototype NL95, 2) BR1, BR8 and prototype SP and 3) prototype FI (Fig 5.2).

Individual proteins were clustered into phylogenetic trees (Fig 5.4). In some cases, phylogenetic protein trees formed more subclusters or branches that the nucleotide trees. For example, *Levivirus* group II lysis protein formed a separate branch (strains T72, KU1) whereas the remaining genogroup II strains (DL20, TL2, DL10, GA) formed a branch off of the group I strains (Fig 5.4). Protein trees generated for *Allolevivirus* were similar to nucleotide phylogenetic clustering. Group III formed two branches, MX1 with M11 and QB-like (6 strains). Group IV generated four subclusters for each protein

tree with the least amount of variation, >90% similarity, noted for the capsid protein.

SimPlots provide a visual picture of regions of similarity when two or more strains are compared. When nucleotide genomes were compared for all Groups I and II strains, SimPlots showed that the replicase was most similar whereas the maturation gene shared the least amount of nucleotide regions (Fig 5.3A). When complete genome nucleotide sequences were compared between groups III and IV, SimPlot graphs showed similar nucleotide regions in the capsid and the 5' portion of the replicase, but least similar in the maturation and 3' region of the replicase (Fig 5.3B)

Table 5.5 Comparison of genomic traits. ***Allolevivirus*** **and** ***Levivirus*****.**

	Allolevivirus	Levivirus
Genome size (nt)	4215-4276	3458-3575
Gene number	4	4
Proteins	4	4
Number of ORF initiation sites	3	4
Protein names	assembly/maturation capsid read-through replicase	assembly/maturation capsid lysis replicase
ORF initiation sites	same for 2nd and 3rd genes	no sharing of ORF initiation sites
replicase gene	single gene	Start codon occurs within lysis protein
3' terminus	5' TCCTCCCA 3'	5' ACCACCCA 3'

.

Figure 5.4 (A) Phylogenetic analysis of each protein: Levivirus I, II.

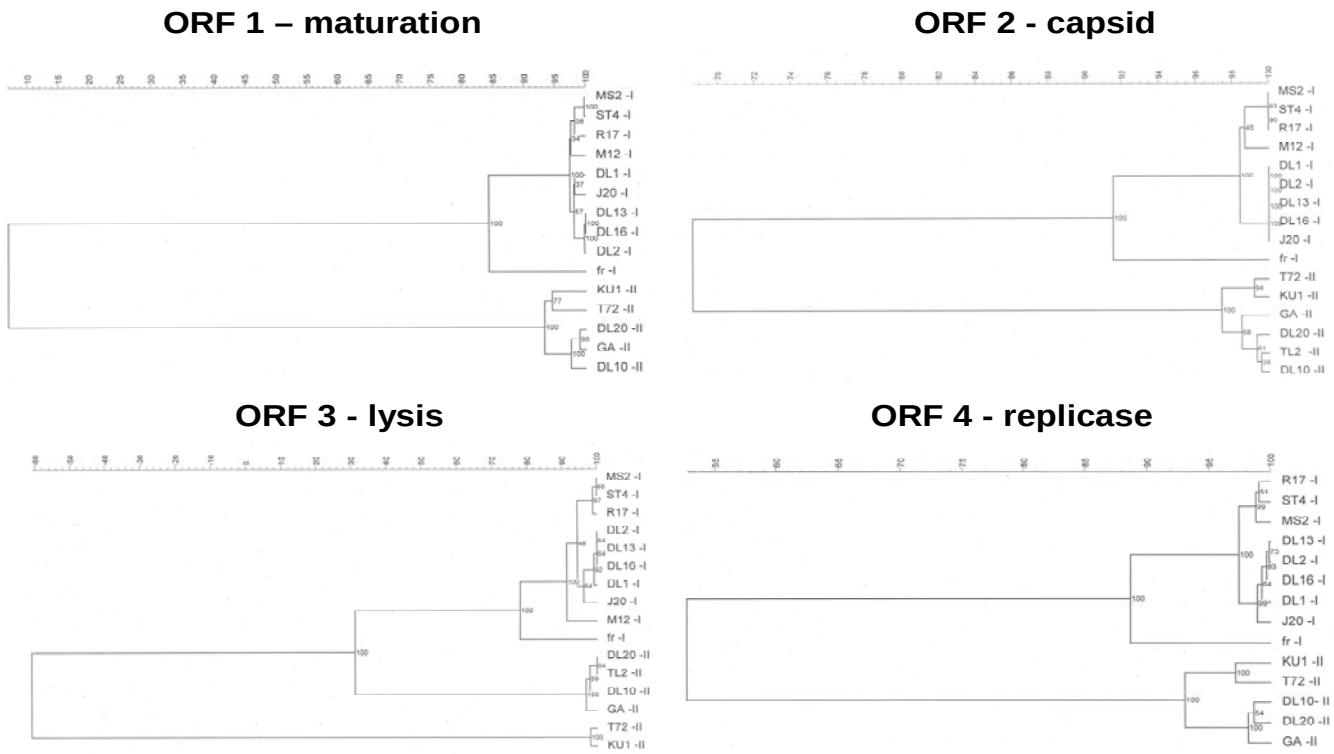
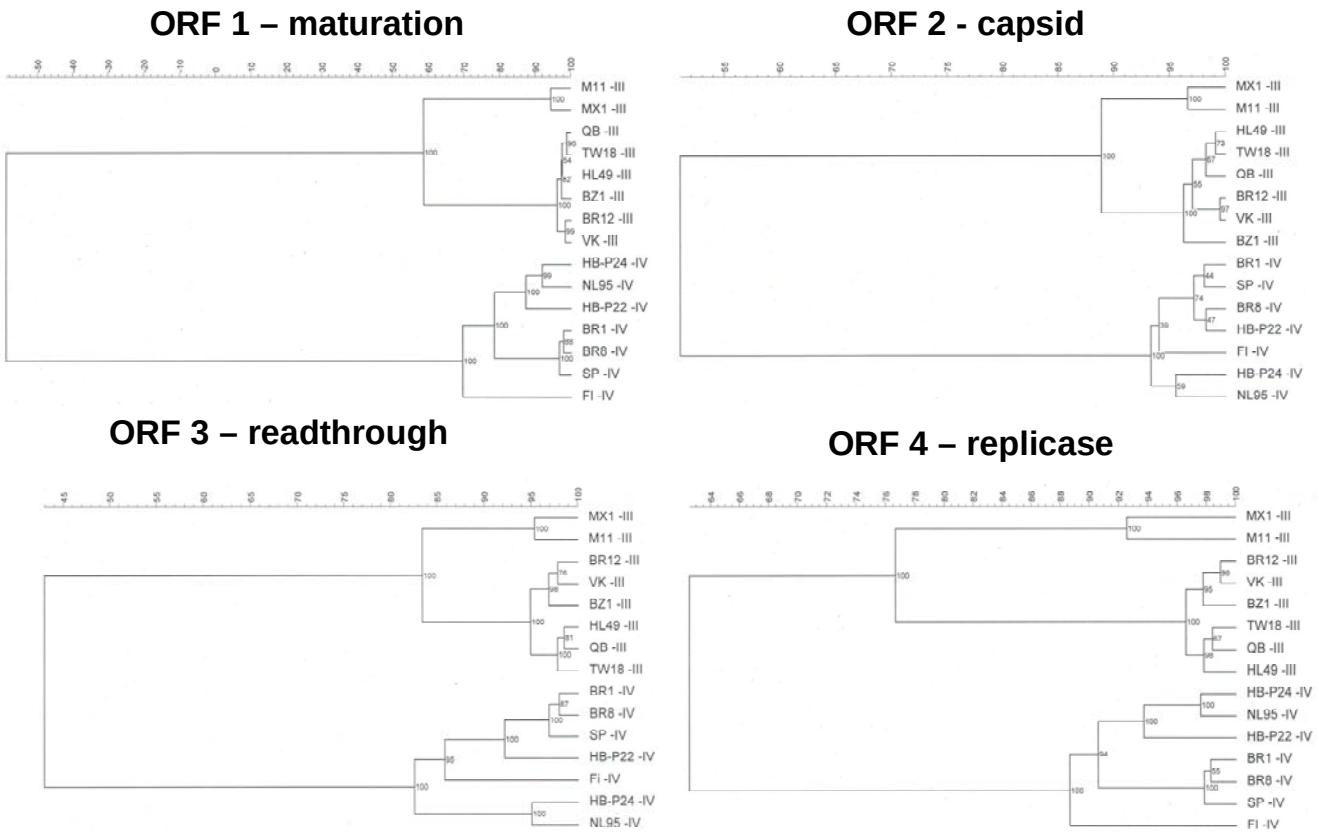


Figure 5.4 (B) Phylogenetic analysis of each protein: Allolevivirus III, IV.



Discussion

Geographically dispersed strains showed more than 90% sequence identity (van Duin, 1998). Our data supports this finding (Fig 5.1, Table 5.2). Although strains DL13, DL2 and DL16 were collected from widely different geographic locations (Table 7.1), sequence alignment among these three strains revealed only 4 nt single-point mismatches throughout the genome with a >99% sequence similarity (Appendix A). With the exception of strain fr, group I sequence identities were greater than 90% (Table 5.2). Strain fr shared a sequence identity no closer than 77.09 % to any of the other group I phages (Table 5.2).

Sequence analyses of the 11 FRNA GenBank strains showed uniformity throughout the *Leviviridae* family (Klovins et al., 2002). Data reported here supports and advances those findings. Capsid/coat protein amino acid length was highly conserved in the *Leviviridae* family ranging from 130-133aa (Table 5.4). Capsid amino acid length is apparently constant in ssRNA phages (Klovins et al., 2002) and may be related to restrictions by capsid infrastructure (Nishihara et al., 2006). In both the previously sequenced *Levivirus* groups I and II phages and in those sequenced in the present study the lysis gene was embedded out of frame as the initiation codon lies at the 3' end of the capsid gene and the termination codon was at the 5' region of the replicase gene. Subsequently, the replicase ORF4 initiation site begins at the 3' region of the lysis gene (Fig 5.3 map, Appendix A). A replicase secondary structure was reported as a conserved amino acid motif, YGDD, in all ssRNA replicases (Olsthoorn et al., 1995). Replicase amino acids from the 30 *Leviviridae* strains reported here and the bacteriophages PP7, AP205 and PRR1 shared the YGDD motif. Also, noteworthy from the sequence data was confirming the *Leviviridae* family conserved

sequence, $CCCA_{OH}$, at the 3' end of the genome (Klovins et al., 2002; van Duin and Tsareva, 2006). This observation was extended by identifying genus-specific *Allolevivirus* and *Levivirus* signatures of 8 nucleotides at the 3' termini that clearly distinguish *Allolevivirus*, 5' TCCTCCCA 3' from *Levivirus*, 5' ACCACCCA 3' (Table 5.5).

In all members of the *Leviviridae* family, the ORF positions for the coliphage genes were preceded by Shine-Dalgarno sequences located 5-16 nt upstream from the start codon(s) (Table 5.3). In prokaryotes, greater than 80% of Shine-Dalgarno sequences occur within 5-13 bases upstream (Ma et al., 2002). Thus, the positions reported here for these viral genomes are similar to the prokaryote criteria.

Stop codons, UAG, UAA, UGA serve as signals for peptide chain termination. During translation of the viral RNA coat protein cistron, the UGA stop codon can be read through, resulting in an additional translated product (Weiner and Weber, 1973). In *Allolevivirus*, a read-through protein resulted when a leaky UGA stop codon is misread as a tryptophan codon (UGG) (van Duin, 2006), influencing regulatory control and efficiency of gene expression (Beier and Grimm, 2001). Alignment of group III nucleotides revealed that the *Allolevivirus* maturation protein stop codon is also a UGA; however, in this instance, a non-leaky codon. This may occur because the 5' and 3' codons flanking the UGA influence translation termination efficiency (Namy et al., 2001, Bertram et al., 2001, Skuzeski et al., 1990). It was proposed that programmed read-through in strain Qβ was regulated by the 3' nucleotides, specifically an A residue, flanking the stop codon in strain Qβ (Engelberg-Kulka, 1981). However, alignment of group III and IV nucleotide data in the present study did not reveal a 3' flanking pattern following the UGA stop codon nor was a 3' residue

observed immediately following the stop codon in all group III strains. Noticeably, Qβ-like strains did contain the 3' A residue but not MX1 or MX11. Interestingly, in all *Allolevivirus* sequences, a 5' pattern emerged at the read-through stop codon but was absent in the maturation UGA stop codon. Beginning 12 nt upstream from read-through UGA, the sequence AAY CCR GCR UAY STOP in group III and AAY CCW GCN UAC STOP in group IV was observed. Nucleotides common to both III and IV are underlined. These findings suggest the upstream sequences may reduce translation termination efficiency of the UGA read-through stop codon in *Allolevivirus sp.*

Representation of protein domain, family conservation and discrete amino acid sequence features, or motifs, were observed by Pfam and PROSITE patterns. Pfam domain groupings were more broadly defined. The replicase beta-chain and capsid domains were distributed throughout the *Leviviridae* family, along with a few additional bacteriophages. In contrast, the lysis and read-through domains were genogroup-specific.

As *Pseudomonas aeruginosa* ssRNA phage PP7 shared secondary regulatory RNA structures with *Leviviridae* it was classified into the genus *Levivirus* (van Duin and Tsareva, 2006) despite the lack of sequence similarity (Olsthoorn et al., 1995, van Duin and Tsareva, 2006) and amino acid clustering (Ruokoranta et al., 2006). Pfam protein domain profile in this study supported the observations that the lysis and capsid proteins of phage PP7 failed to cluster to these same proteins in members of the *Levivirus* or the capsid and read-through proteins in members of the *Allolevivirus*. Data from the present study also indicated that phage PP7 replicase protein did not cluster to the replicase members of the *Levivirus* but did cluster to *Allolevivirus* replicase protein. Common to ssRNA phages, the replicase amino

acid motif, YGDD, was present in phage PP7 replicase. Phage PP7 maturation protein did cluster to this same protein in both *Levivirus* and *Allolevivirus*. Data from this and previous studies suggest that the ssRNA phage PP7 genetic map was structured similar to *Levivirus* groups I and II (Ruokoranta et al., 2006).

Phage PRR1 has a broad host range related to a plasmid IncP-encoded receptor, adsorbs to host pili and displays a genetic map similar to *Leviviridae* (Ruokoranta et al., 2006). PRR1 shared approximately 43-48% sequence identity to other ssRNA phages and clustered outside the *Levivirus* and *Allolevivirus* genera (Ruokoranta et al., 2006). PRR1 genetic map was similar to *Leviviridae* and subsequently phage PRR1 was grouped into the *Levivirus* genus (Ruokoranta et al., 2006). Pfam analysis presented here resulted in phage PRR1 sharing Pfam domains with *Leviviridae* maturation, capsid and replicase proteins. PRR1 displayed the ssRNA amino acid YGDD replicase motif but did not share the signature *Levivirus* 3' terminus ACCACCCA.

The ssRNA phage from *Acinetobacter*, AP205, shared Pfam domains in the *Levivirus* and *Allolevivirus* maturation and replicase proteins only. Amino acid compositions from AP205 and their corresponding coat, maturation, lysis and replicase proteins clustered outside the *Levivirus* and *Allolevivirus* tree (Ruokoranta et al., 2006). AP205 lacked significant sequence similarity but shared important structural features with *Leviviridae* (Klovins et al., 2002). As with phages PRR1 and PP7, phage AP205 did not share the 3' termini *Levivirus* or *Allolevivirus* signatures but had the replicase YGDD motif. NCBI GenBank classified bacteriophages PRR1, PP7 and AP205 into an “unclassified” category.

Using PROSITE, conserved amino acids and subsequent motifs identified in amino

acid sequences of the phage proteins provided insight to structural features. A cell-attachment motif (RGD) was identified in both the maturation and/or read-through proteins in the majority of *Allolevivirus* strains but was absent in *Levivirus* strains. The function of the RGD motif in *Leviviridae* coliphages has yet to be experimentally demonstrated but may be explained because in *Levivirus* strains the phage attaches to the host's pili via the maturation protein; in *Allolevivirus* strains both the maturation and read-through proteins are required for phage infection (van Duin, 1999). The Arg-Gly-Asp (RGD) motif was shown to be involved in cell-to-cell adhesion in the passaged foot-and-mouth disease virus (Martinez et al., 1997), in enterovirus, echovirus 9 strain Barty, coxsackievirus A9, echovirus 22 (Nelsen-Salz et al., 1999) and the blue-tongue virus (Tan et al., 2001). In nearly all *Astrovirus sp.* an RGD or similar integrin-recognition motif was observed (van Hemert et al., 2007). In contrast, a second RGD motif in a SAT1 foot-and-mouth virus was not necessary for cell-to-cell attachment (Storey et al., 2007).

As more Gram negative bacteriophage sequences are elucidated, the protein domains, phylogenetic relationships and novel virus groups will likely emerge and enrich the database.

Conclusion

The findings of this study agree with previously determined FRNA features and phylogenetic analyses which concluded that the *Leviviridae* contain two genera and four distinct genogroups consisting of two genogroups within each genus. Despite the fact that genogroup I strain fr, and genogroup III strains MX1and M11 only shared between a 70-78% sequence identity with the rest of the strains in their respective genogroups, the analyses suggested that fr does belong in *Levivirus* group I, and MX1 and M11 belong in *Allolevivirus*

group III. Distinguished features such as amino acid and nucleotide similarity and catalytic domain location tend to sub-cluster the strains. For example, Qβ-like strains clustered together when compared to MX1 and M11 in group III. Genogroups within each genera shared approximately 50% sequence identity whereas between the two genera, *Levivirus* and *Allolevirus*, <40% nucleotide sequence identity was observed. Genome organization, amino acid conservation and identical or very similar nucleotide start and stop positions supported the genogroup designation. In addition, eight nucleotides on the 3' termini clearly distinguish the *Allolevivirus*, 5' TCCTCCCA 3', from the *Levivirus*, 5' ACCACCCA 3'.

Summary

- Alignment of nine group I environmental strains with reference strain MS2 showed identical ORF and start codon positions for all four *Levivirus* genes, indicating that the sequence data generated in the present study was valid.

- All strains for groups II, III and IV had similar results in that the sequenced genomes and gene maps showed identical or very similar nucleotide positions to the GenBank reference strains.

- Basic Local Alignment Search Tool (BLAST) of all environmental strains placed them into their respective genogroups.

- Amino acid composition was similar among FRNA coliphage genogroups, further validating the nucleotide sequences and the groupings based upon them.

References

Beier, H, and M Grimm. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucl Acids Res, 29(23):4767-4782.

Bertram G, S Innes, O Minella, JP Richardson, and I Stansfield. 2001. Endless possibilities: translation termination and stop codon recognition. Microbiol, 147:255-269.

Bollback, J.P. and J.P. Huelsenbeck. 2001. Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family *Leviviridae*). J Mol Evol 52:117-128.

Buchen-Osmond, C. (Ed.) 2004. Levivirus. In ICTVdB - *The Universal Virus Database,* version 3. ICTVdB Management, Columbia University. New York, NY.

Crawford, EM, and RF Gesteland. 1964. The adsorption of bacteriophage R17. Virol, 22:165-167.

EPA. 2004. Quality assurance/quality control guidance for laboratories performing PCR analyses on environmental samples. EPA 815-B-04-001. October 2004.

Engelberg-Kulka, H. 1981. UGA suppression by normal tRNA ^{Trp} in *Escherichia coli*: codon context effects. Nucl Acids Res, 9(4):983-991.

Fiers W, R Contreras, F Duerinck, G Haegeman, D Iserentant, J Merregaert, W Min Jou, F Molemans, A Raeymaekers, A Van den Berghe, G Volckaert and M Ysebaert. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature, 260:500-507.

Grabow, W.O.K., 2001. Bacteriophages: update on applications as models for viruses in water. Water S.A. 27, 251-268.

Havelaar**,** A. H., M. van Olphen, and Y. C. Drost**.** 1993. F-specific RNA bacteriophages are adequate model organisms for enteric viruses in fresh water. Appl. Environ. Microbiol. 59**:**2956–2962.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules, p. 21-123. In H. N. Munro (ed.), Mammalian Protein Metabolism. Academic Press, New York.

Karnik, S. and M. Billeter. 1983. The lysis function of RNA bacteriophage Qβ is mediated by the maturation protein. EMBO J. 2:1521-1526.

Klovins, J., G.P. Overbeek, S.H.E. van den Worm, H.-W. Ackermann and J. van Duin. 2002.

Nucleotide sequence of a ssRNA phage from Acinetobacter: kinship to coliphages. J Gen Virol, 83, 1523-1533.

Loeb, T. and N.D. Zinder. 1961. A bacteriophage containing RNA. Proc. Natl. Acad. Sci. U.S. 47:282-289.

Lole, KS., RC Bollinger, RS Paranjape, D Gadkari, SS Kulkarni, NG Novak, R Ingersoll, HW Sheppard and SC Ray. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype c-infected serconverters in India, with evidence of intersubtype recombination. J Virol, 73(1):152-160.

Ma J, A Campbell and S Karlin. 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J Bact 184(20):5733-5745.

Martinez, MA, N Verdaguer, MG Mateu, and E Domingo. 1997. Evolution subverting essentiality: dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus. Proc Natl Acad Sci, 94:6798-6802.

Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (1995) Virus taxonomy: The classification and nomenclature of viruses. The sixth report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Vienna

Namy, O, I Hatin and JP Rousset. 2001. Impact of the six nucleotides downstream of the stop codon on translation termination. EMBO Reports, 2(91):787-793.

Nelsen-Salz, B, HJ Eggers, and H Zimmermann. 1999. Integrin $\alpha_v\beta_3$ (vitronectin receptor) is a candidate receptor for the virulent echovirus 9 strain Barty. J Gen Virol, 80:2311-2313.

Nishihara T., S. Fujisaki, Y. Nishimura, Y. Minami and T. Yubisui. 2006. Analysis of six new genes encoding lysis proteins and coat proteins in *Escherichia coli* Group A RNA phages. Microbiol. Immunol., 50(1), 61-66.

Olsthoorn, RCL, G Garde, T Dayhuff, JF Atkins, and J van Duin. 1995. Nucleotide sequence of a single-stranded RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory RNA structures. Virol, 206:611-625.

Paranchych, W. 1975. Attachment, ejection and penetration stages of the RNA phage infectious process. In: ND Zinder (ed). RNA Phages. Cold Spring Harbor Laboratory. pp. 85-111.

Ruokoranta, TM, AM Grahn, JJ Ravantti, MM Poranen and DH Bamford. 2006. Complete genome sequence of the broad host range single-stranded RNA phage PRR1 places it

in the *Levivirus* genus with characteristics shared with Alloleviviruses. J Virol, 80(18):9326-9330.

Skuzeski, JM, LM Nichols and RF Gesteland. 1990. Analysis of leaky viral translation termination codons in vivo by transient expression of improved beta-glucuronidase vectors. Plant Mol Biol, 15(1):65-79.

Stewart, J.R., J. Vinjé, S.J.G. Oudejans and G.I. Scott., 2006. Sequence variation among group III F-specific RNA coliphages from water samples and swine lagoons. Appl. Env. Micro, 72, 1226-1230.

Storey, P, J Theron, FF Maree and HG O'Neill. 2007. A second RGD motif in the 1D capsid protein of a SAT1 type foot-and-mouth disease virus field isolate is not essential for attachment to target cells. Viral Res, 124:184-192.

Sundram A, N Jumanial, and MM Ehlers. 2006. Genotyping of F-RNA coliphages isolated from wastewater and river water samples. Water SA, 32(1):65-70.

Tan, BH, E Nason, N Staeuber, W Jiang, K Monastryrskaya and P Roy. 2001. RGD tripeptide of bluetongue virus VP7 protein is responsible for core attachment to *Culicoides* cell. J Virol, 75(8):3937-3947.

van Duin, J. 1988. The single-stranded RNA bacteriophages. In: Fraenkel Conrat H and Wagner RR. (eds). *The Bacteriophages*. Series *The Viruses*. Chapter 4. Plenum Press, NY. pp. 117-167.

van Duin, J., 1998. The single-stranded RNA bacteriophages. In: Fraenkel Conrat, H., Wagner, R.R. (eds). The Bacteriophages. Series *The Viruses*. Plenum Press, NY. pp. 117-167.

van Duin, J. 1999. Single-stranded RNA phages (*Leviviridae*). In: Granoff, A and R Webster (eds). Encyclopedia of Virology. Academic Press, London. pp. 1663-1668.

van Duin, J. 2000. Family *Leviviridae*. In: M.H.V. van Regenmortel, C.M. Fauquet, D.H.L. Bishop, E.B. Carstens, M.K. Estes, S.M. Lemon, J. Maniloff, M.A. Mayo, D.J. McGeoch, C.R. Pringle, R.B. Wickner (eds). Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego. pp. 645-646.

van Duin, J, and Tsareva, 2006. Single-stranded RNA phages. In: *The Bacteriophages* by S T Abedon. R L Calendar, ed. Oxford University Press. pp:175-196.

Weber, K, and W Konigsberg. 1975. Proteins of the RNA phages. In: ND Zinder (ed).

RNA Phages. Cold Spring Harbor. pp. 51-84.

Weiner, AM and K Weber. 1973. A single UGA codon functions as a natural termination signal in the coliphage Qβ coat protein cistron. J Mol Biol, 80:837-855.

van Hemert, FJ, VV Lukashow and B Berkhout. 2007. Different rates of (non-) synonymous mutations in astrovirus genes; correlation with gene function. Virol J, 4:25-37.

Vinjé, J., S.J.G. Oudejans, J.R. Stewart, M.D. Sobsey and S.C. Long., 2004. Molecular detection and genotyping of male-specific coliphages by reverse-transcription-PCR and reverse line blot hybridization. Appl. Env. Micro., 70(10), 5996-6004.

VI. Genomic sequences of two novel *Levivirus* FRNA coliphages (*Leviviridae):* Evidence for recombination?

Abstract

Male-specific ssRNA coliphages, family *Leviviridae,* contain two genera and four distinct genogroups. The *Levivirus* genus is subdivided into genogroups I and II whereas the *Allolevivirus* genus consists of genogroups III and IV. During an environmental genotyping study of *Leviviridae* ssRNA coliphages, two novel strains became evident. Nucleotide sequences and phylogenetic analysis of a 189 bp amplicon in the replicase gene clustered the strains between *Levivirus* genogroups I and II, leading to a proposed genogroup I *Levivirus* subcluster, termed JS-like. Seventeen strains from genogroups I and II, including two JS strains, were used to examine the genomic and phylogenetic relationships among these *Levivirus* groups. The two JS strains were 96.73% similar in nucleotide sequences to each other and 80-84 % nucleotide sequence similarity was shared between the JS strains and nine MS2-like strains. Amino acid composition between JS strains and MS-2 like strains of the maturation, capsid and lysis proteins shared 99-100%, 98-100% and 95-100%, respectively. However, the replicase amino acid sequences of the JS strains shared only 84-85% amino acid similarity to nine MS2-like strains. *Levivirus* JS strains diverged from group I in the replicase gene downstream from the conserved catalytic site. Analysis of complete genome sequences, amino acid composition, phylogenetic relationships and unique clustering suggest the JS strains are recombinants.

Introduction

Vinjé et al (2004) investigated the genetic diversity of male-specific (FRNA) phages. In that Vinjé study, a phylogenetic analysis of 32 *Levivirus (*genogroups I and II) field strains using a 189 bp replicase gene fragment revealed three main clusters: genogroup I, genogroup II and a potential novel group, designated JS, which clustered between genogroup I and genogroup II. The putative JS group, represented by phages, WWTP1_50 and 2GI13, had a >40% sequence diversity in the 189 bp replicase gene sequence with strains from genogroups I and II. As these strains were isolated from geographically different regions (Massachusetts and South Carolina) Vinjé et al (2004) proposed that JS may form a stable lineage and suggested that further genomic sequence and serological data were needed to confirm whether these strains form a novel genogroup or whether these strains are the result of recombination or rearrangement events (Chetverin, 1999). In the present study complete genome sequences of two additional phages, belonging to the putative JS group (Sobsey et al., 2006; Love et al., 2008), were determined allowing a phylogenetic study into the nature of this proposed subgroup.

Whether or not the putative JS subgroup represents a novel genogroup, recombinations (sequence exchange and rearrangements) between RNA molecules may have occurred in these viruses. Largely responsible for the diversity of RNA viruses (Lai, 1992) RNA-RNA recombination has been shown to occur in several positive-sense, ssRNA human and animal viral taxa including caliciviruses, coronaviruses, hepatitis, dengue, enteroviruses and astroviruses (Cristina and Colina, 2006; Pantin-Jackwood et al., 2006; Holmes et al., 1999;

Oberste et al., 2004; Banner and Lai, 1991; Oprisan et al., 2002; Walter et al., 2001; Jiang et al., 1999; Belliot et al., 1997).

RNA recombination events occur, in some cases, when two or more strains infect the same host. Proposed models for the formation of novel RNA sequences include (i) cleavage and ligation in RNA molecules or RNA secondary structures (Lutay et al., 2007), (ii) replicative template switching whereby the RNA-dependent RNA polymerase (replicase) switches from one template to another RNA template, also known as copy choice (Chetverin, 1999; Chetverin et al., 2005), and (iii) RNA transesterification reaction that occurs when the polymerase adds a separate RNA fragment to the 3' terminus of the original RNA template (Chetverin, 1999)

The first indication of non- replicative RNA recombination in a male-specific FRNA phage was reported by Munishkin et al., (1988) who found small, nonhomologous, recombinant RNA molecules produced *in vitro* in a purified template-free Qβ replicase molecule. These investigators noted similar RNA molecules were present in *E. coli* cells infected with phage Qβ. Chetverin et al (2005), studied this phenomenon by observing the formation of novel sequences in a colony of RNA molecules, suggested that this recombination event occurred as a transesterification reaction catalyzed by a conformation acquired by Qβ replicase during RNA synthesis (Chetverin et al., 2005; Chetverin, 1999). Nucleotide sequences of recombined RNA molecules were non-homologous to the parent RNA and were formed in the absence of DNA intermediates, demonstrating an RNA

recombination mechanism in the presence of Qβ replicase (Chetverin, 1999). Therefore, it is plausible to have recombination in environmental ssRNA male-specific coliphage (*Leviviridae*) isolates.

Two JS strains, DL52 and DL54, were isolated during an environmental genotyping study of *Leviviridae* FRNA phages (Sobsey et al., 2006; Love et al., 2008). These phages were placed into the putative JS subgroup using the genotyping methods of Vinjé et al (2004). The objective of the present study was to determine whether the existence of a novel JS-like subgroup representing a third *Levivirus* cluster as proposed by Vinjé et al (2004) could be verified. The approach was to analyze recently generated nucleotide and amino acid sequence data from entire genomes of 10 levivirus group I strains, 5 levivirus group II strains and two JS group strains. Analysis of the novel JS strains should provide evidence as to whether or not these *Levivirus* strains were genogroup I, II, a combination of groups I and II (recombinant strain) or a unique genogroup. To further understand the composition of these JS strains, complete genomic sequencing, amino acid composition and phylogenetic analyses were examined.

Purpose

- Evaluate the genomic nucleotide and amino acid sequences to determine the uniqueness of JS strains and to clarify why JS strains do not hybridize to group I or group II reverse-line blot hybridization probes.

Approach

- Sequence the two JS-like strains, DL52 and DL54, to determine their taxonomic classification.

- Map Open Reading Frames and determine gene locations.

- Translate nucleotide sequences into amino acid compositions for each protein.

- Compare JS strains to *Levivirus* groups I and II.

- Analyze similarities and differences between JS and groups I and II using various bioinformatic methods.

Materials and Methods

Coliphage Isolates and Propagation

FRNA phage strains CICEET 29 and CICEET 24 were isolated and placed into the putative JS subgroup by Love et al., (2008; Sobsey et al., 2006) using the genotyping methods of (Vinjé et al., 2004). CICEET 29, renamed DL52, was isolated from estuarine waters in Rachel Carson W Reserve (Beaufort), NC, and CICEET 24, renamed DL54, and was isolated from Naragansett Bay, RI.

Sequencing was performed on plaque-purified coliphages which were enriched using *Escherichia coli* HS(pFamp)R as host (Vinjé et al., 2004). Overnight enrichments were centrifuged (3,220 X g for 10 min) to pellet host cells and debris. A 1:1 (V/V) chloroform to virus supernatant mixture was vortex mixed and centrifuged again. Approximately 1-2 ml aliquots of the supernatant containing the phage enrichments were stored frozen at −75 ^{o} C.

Coliphage titers were determined using a single agar layer procedure (SAL) (US EPA Method 1602, 2001). The procedure was as follows. A 1 ml of overnight *E. coli* Famp host culture was transferred into 50 ml trypticase soy broth with streptomycin and ampicillin (TSB/strep/amp) and grown 4 hr to log phase. A 150 ml volume of trypticase soy agar, TSA (4.5 g TSB and 1.2 g agar), was autoclaved and placed into a water bath (47-55 ^{o} C). A 300 ul aliquot of 500X strep/amp was added into the cooled 150 ml TSA. Serial 10-fold dilutions of the purified virus were prepared. For 10-fold stock dilution, 100 ul of stock was transferred into 900 ul phosphate buffered saline (PBS), mixed and this process repeated to a serial dilution 10^{-15}. To sterile15 ml plastic tubes, labeled -1 to -15, 1 ml of the 4 hr *E. coli*, 100 ul of the appropriate virus dilution and 10 ml TSA were added. After quick manual mixing the

tube contents were poured into labeled 20mm petri plates which were then cooled, allowed to dry, inverted and incubated overnight at 37 ^{o} C. The coliphage titer was determined by counting the plates having a minum of 5 to a maximum of 20 plaques per plate. Titer concentration was reported as PFU/ml.

Coliphage RNA Isolation

Coliphage RNA was extracted from the purified viral supernatant using a QIAamp viral RNA mini kit (Qiagen, Valencia, CA) as follows. A 200 ul volume of purified virus stock was added into the 25 ul protease tube. Into the same tube, 200 ul of carrier RNA/Buffer AL mixture was added and the combined solutions were vortex mixed, incubated 56 ^{o} C for 15 minute and clarified by a 1 minute centrifugation. To the same tube, 250 ul ethanol (EtOH) was added, vortex mixed and incubated 5 min at room temperature (RT). The lysate was applied into the QIAamp spin column and centrifuged 3-4 min at 6000 xg (8000 rpm) and the contents of the collection tube was discarded. To the column, 500 ul of Buffer AW2 was added, centrigued 1 min and the contents of the collection tube was discarded. To the column, 500 ul EtOH was added, centrifuged 1 min and collection tube was discarded. The column was transferred to a clean collection tube and centrifuged for 1 minute. The column was again transfered a new, nuclease-free tube and 50 ul RNase free water was added onto column and incubated 5 min at RT. The final centrifugation, (14,000 rpm) for 1 min and the recovered purified RNA was frozen at −20 ^{o}C.

Generating cDNA from Polyadenylated RNA

Phage MS2 serves as a positive, procedural control. Viral RNA was 3' polyadenylated with yeast PolymeraseA and 25 mM ATP in a 50 ul reaction volume (USB, Inc, Cleveland,

OH). The reaction was prepared with 10 ul 5X Reaction Buffer, 10 ul RNA, 2 ul 25 mM ATP, 0.7 ul 600 U Poly(A)Polymerase and 27.3 ul nuclease-free water. The mixture was incubated at 37^0 C for 5 min and placed on ice for enzymatic termination. Polyadenylated RNA was either immediately frozen or used as a template for cDNA.

Full-length cDNA was prepared using oligo-dT reverse primer supplied with the reverse transcriptase MonsterScript 1st Strand cDNA Synthesis Kit (EpiCentre, Madison, WI) or with a gene-specific reverse primer. To a 250 ul thin-walled PCR tube, the following reagents were added: 4 ul nuclease-free water, 10 ul polyadenylated RNA or RNA template and 1 ul of 10 uM PolyT primer or 1 ul of 10uM gene-specific primer. The mixture was heated for 1 min at 65 ^{o} C and chilled for 1 min on ice. To the same tube, 1 ul MonsterScript Reverse Transcriptase and 4 ul of 5X cDNA reaction buffer were added. The mixture was placed in a thermocycler with the following cycle regime: 37 ^{o} C for 5 min, 42 ^{o} C for 5 min, 60 ^{o} C for 40 min and 90 ^{o} C for 5 min and chilled on ice for 1 min. The single-stranded cDNA was either frozen or used for PCR template.

5' Amplification of cDNA Ends

The nucleotide sequence of the 5' region was determined by rapid amplification of cDNA end (RACE) with the Smart Race cDNA Amplification Kit (Clontech, Mountain View, CA). First-strand cDNA synthesis was prepared on ice in a 250 ul thin-walled PCR tube by combining 3 ul RNA, 1 ul of 10uM gene-specific reverse primer and 1 ul Smart oligo (from kit). The 5 ul reaction volume was briefly centrifuged and the following components were added: 2 ul of 5X First Strand buffer, 1 ul of 20 mM DTT, 1 ul of 10 mM dNTP and 1 ul SuperScript II (Invitrogen, Carlsbad, CA). Following a brief centrifugation, the mixture was

incubated for 90 min at 42 ^{o} C. To dilute the first-strand cDNA, 20 ul of Tricine-EDTA buffer was added and heated for 7 min at 72 ^{o} C. The reaction generated double stranded cDNA. The cDNA was frozen at - 20 ^{o} C and used for subsequent PCR reactions. Concentration of cDNA was determined with a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE).

PCR, Cloning and Sequencing

The cDNA was amplified by using Phusion DNA Polymerase (New England Biolabs, Ipswich, MA), with final concentrations of 1X of 5X Phusion Buffer, 0.2 mM dNTP, 1 ul of 10 uM forward primer, 1 ul of 10uM reverse primer, 3% DMSO, 2 ul cDNA and 0.5 ul Phusion Taq in a 50 ul reaction using the following cycle parameters: one round denaturation at 98 ^{o}C (1 min), 35 rounds at 98 ^{o}C (30 sec), 48 ^{o}C (1 min), 72 ^{o}C (3 min) followed by 10 min extension at 72 ^{o}C. For each reaction, positive controls were prepared using primers MJV82 and JV81 for *Leviviruses* and MJV82 and JV41 with *Allolevivirus*es. A no-template negative control was included.

To verify the generation of full-length cDNA, the 5' end of phage MS2 was PCR amplified using primers MS25 and MS23 and the same protocol as stated above (Lovmar *et al.*, 2003).

PCR products were separated by electrophoresis in a 1.5% agarose gel, stained with SYBR Gold Nucleic Acid Gel Stain (Molecular Probes, Carlsbad, CA) and visualized under blue light (Dark Reader Transilluminators, Clare Chemical Research, Dolores, CO).

Blunt-end PCR products were excised using a gel extraction tool (USA Scientific Plastics, Ocala, FL) and then purified (QuickClean 5M Gel Extraction Kit, GenScript Corporation, Piscataway, NJ).

Excised bands were weighed in 1.5 ml microcentrifuge tubes and 3 volumes of Binding Solution II per gel slice were added. The gel solution was heated at 50 ^{o} C until melted. One volume of isopropanol was added and the mixture was transferred into the Genprep spin column and centrifuged for 1 min at 12,000 rpm. The column effluent was discarded and 500 ul of Wash Buffer was added, the column was centrifuged and the liquid waste discarded. The 500 ul wash was repeated and waste discarded. The column was placed into a clean 1.5 ml microcentrifuge tube, 30 ul of Elution Buffer was added and the column was incubated 1 min at room temperature. The tube was centrifuged 1 min at 12,000 rpm and the collected DNA eluate was transferred to a clean tube. Concentration of the DNA was determined using a NanoDrop spectrophotometer. The DNA was either cloned or the PCR product was sequenced.

Gel-purified DNA was cloned using a ZeroBlunt TOPO Cloning Kit, pCR-Blunt II TOPO plasmid kit (Invitrogen, Carlsbad, CA). To prepare a TOPO cloning reaction, 4 ul DNA, 1 ul salt solution and 1 ul pCR II Blunt TOPO were added to a nuclease-free tube and incubated 30 min at room temperature. To a vial of One Shot *E. coli* competent cells, 2 ul of the TOPO cloning reaction were added and incubated on ice for 30 min. The cells were heat-shocked at 42 ^{o} C and immediately placed on ice. A 250 ul volume of SOC medium was added to the One Shot cells and shaken (200 rpm) for 1 hr at 37 ^{o} C. Fifty ul of transformed cells were plated onto pre-warmed LB agar plates containing 50 ug/ml kanamycin, the

transformed cells were spread to isolate colonies and the plates were incubated overnight at 37 ^{o} C.

Colonies of transformed *E. coli* cells were screened for positive inserts using whole-cell PCR. Using aseptic techniques, individual colonies were selected with a sterile toothpick, the toothpick was briefly rinsed into a 50 ul Phusion master mix (as described above) then dropped into 10 ml LB/kan broth and incubated overnight at 37 ^{o} C. Whole-cell PCR was performed on individual colonies using Phusion DNA Polymerase with the following cycle modifications: one round denaturation at 98 ^{o}C (3 min), 35 rounds at 98 ^{o}C (10 sec), 57 ^{o}C (30 sec), 72 ^{o}C (30 sec) followed by 10 min extension at 72 ^{o}C. Amplicons were separated by electrophoresis in 1.5% agarose gel in 0.5X Tris-acetate-EDTA (TAE), stained with 20 ug/ml ethidium bromide and visualized under UV light (UVP, Upland, CA). Those clones with the appropriate size PCR amplicon were selected for plasmid purification.

Positive clones were plasmid-purified (QIAprep Spin Miniprep Kit, Qiagen, Valencia, CA). An *E. coli* colony that had been selected with a toothpick and incubated overnight was centrifuged 10 min at 8000 rpm (6800 xg). The supernatant was discarded and the cell pellet was resuspended and processed as follows. A 250 ul volume of Buffer P1 was added to the cell pellet, vortexed to mix and transferred to a clean 1.5 ml microcentrifuge tube. A 350 ul volume of Buffer P2 was added to the resuspended pellet and mixed by inversion followed by addition of 350 ul Buffer N3. The buffer mixture was inverted 4-6 times and centrifuged 10 min at 13,000 rpm (17,900 xg). The supernatant was decanted into the QIAprep spin column and centrifuged 1 min at 13,000 rpm. The column effluent was discarded and the column was washed with 500 ul Buffer PB, the column was centrifuged 1 min and the wash was

discarded. To the spin column, 750 ul Buffer PE was added, the column was centrifuged and the wash was discarded. The column was incubated with 50 ul Buffer EB to elute the plasmid. Following a 1 min incubation at room temperature, the column was centrifuged 1 min. Purified plasmid was shipped frozen for sequencing (Sequetech, Mountain View, CA). Each cDNA PCR amplicon was cloned and sequenced in triplicate to account for errors. In some cases, the PCR product was sequenced directly. To achieve publication quality data, both forward and reverse strands were sequenced (Sequetech, Mountain View, CA). This process was repeated until complete genomes were obtained.

To avoid contamination, a PCR hood (AirClean 600, AirClean Systems, Raleigh, NC) located in a designated room was used to prepare master mixes separate from template additions. PCR amplification, electrophoresis, template and/or viral preparations (EPA, 2004) were conducted in individual assigned rooms based on designated use.

Primer Design

PCR amplicons were first generated from an approximate 200 nt region using degenerate primers (MJV82 and JV81 for *Leviviruses*) specific to the replicase region (Vinjé et al., 2004) and sequenced (UNC, NC). The combination of these primer sets was termed "generic PCR" as the primers were genus specific but not genogroup specific. A strain-specific forward primer was designed from this 200 nt region. A combination of the strain-specific forward primer and a PolyT reverse primer amplified an approximate 1 kb PCR fragment. The amplicon was cloned and sequenced. As sequences of the fragments were generated, reverse primers were designed to amplify overlapping sections of the genome. Forward primers were designed by alignment (BioEdit v7.0.1 Clustal W application; Hall,

1999) of representative group I (J20, M12) and II (KU1) strains sequenced in this study or available in GenBank.

Sequence Analyses and Open Reading Frames

JS sequences were compared to nucleotide and/or amino acid sequences from 10 group I strains (MS2, DL1, DL2, DL13, DL16, ST4, R17, J20, M12, fr) and 5 group II strains (T72, DL10, DL20, GA, KUI). Sequences and/or amino acids were aligned using BioEdit v7.0.1 ClustalW application (Hall, 1999). Basic Local Alignment and Search Tool (BLAST) finds regions of local similarity between sequences and is used to search similar matches in the National Center for Biotechnology Information (NCBI) genetic databse. BLAST analyses for sequence and phylogenetic confirmation were performed on each individual FRNA clone or PCR fragment. Open Reading Frames (ORF) were determined using BioEdit (Hall, 1999).

Nucleotide percent similarity and dendograms were constructed using Bionumerics software v. 3.5 (Applied Maths, Saint-Martens-Latem, Belgium). Phylogenetic trees were built using the global cluster analysis performed on multiple aligned sequences and clustered by UPGMA using the Jukes and Cantor correction. A bootstrap analysis, based on 10,000 substitutions, was used to measure cluster significance.

Cluster analysis of Levivirus Groups I, II and JS phages were generated from pairwise similarities of the amino acid sequences of their replicase genes (Bionumerics). Standard deviations of the average similarites of the clusters were determined using Bionumerics. The resulting phylogenetic tree produces a cophenetic correlation which represents the faithfulness of the clusters expressed on a percentage basis (Bionumerics).

Amino Acid Analysis

Amino acid compositions for each of the four genes were determined using a computer-generated DNA-to-protein translation tool, ExPASY (http://ca.expasy.org/). Prediction of protein sequence motifs were identified by PROSITE (http://ca.expasy.org/).

Amino acid sequence data were analyzed using BioNumerics Software v.3.5 (Applied Maths, Saint-Martens-Latem, Belgium). Phylogenetic trees were built by global cluster analysis performed on multiple aligned sequences and clustered by UPGMA using the Jukes and Cantor correction (Jukes & Cantor, 1969). Cophenetic correlations and cluster Cutoff method were employed to measure faithfulness and relevancy of the clusters. Average similarities with standard deviations were calculated for the relevant clusters.

Relationships among aligned amino acid sequences were depicted in similarity plots generated by SimPlot, v3.5.1 (Lole et al., 1999). The SimPlot program determines the percent identity between a reference and the queried sequence. Percent similarity was calculated within a sliding window of 160 bp wide with a step size between plots of 10 bp.

Results

Degenerate replicase primers in an RT-PCR assay identified the two strains, DL52 and DL54, to the *Levivirus* genus. Nucleotide sequences of the 189 bp amplicon classified the strains as JS-like (Vinjé et al., 2004). Reverse-line blot hybridization failed to genotype the two strains into subgroups I or II.

A total of 17 strains (MS2, ST4, DL1, DL2, DL13, DL16, R17, M12, J20 and fr in genogroup I, DL52 and DL54 in the JS genogroup and T72, DL10, DL20, GA and KU1 from

genogroup II) were used to examine the relationships among nucleotides and amino acids in the *Levivirus* genus. The first 9 strains in genogroup I are referred to as "MS2-like."

With respect to nucleotide sequences, the MS2-like strains shared 91.68-99% full-length genome nucleotide sequence similarity to each other (Fig 5.2) and the two JS strains, DL52 and DL54, were 96.73% similar in nucleotide sequences to each other (Table 6.2). The JS strains were more similar to MS2-like genogroup I FRNA coliphage (80-84 % sequence similarity) than the fr strain was to MS2-like phages (75.27-77.65% sequence identity). Despite their sequence similarities, genome lengths for JS strains (3525 nt) were shorter than the MS2-like group I (3569-3575 nt) (Table 6.1) but longer than genogroup II genomes (3458-3486 nt) (Table 5.1). Numerous deletions in the 3' untranslated region and a portion of ORF4 (replicase) in JS strains accounted for the decreased length (Appendix C).

Within genogroup I, the amino acid sequences of all four proteins of strain fr were distinctly different from the proteins of the MS2-like strains (Fig 5.4A, Fig 6.2). A different pattern, however, emerged when comparing the sequences of the four proteins of the JS strains to the protein sequences of the MS2-like genogroup I strains. Sequence similarities of the maturation, capsid and lysis proteins of the JS strains were very similar to those of the MS2-like group strains, sharing 99-100%, 98-100% and 95-100% sequence similarities, respectively (Table 6.3, Fig 6.1). However, the replicase protein sequences of the JS strains were quite dissimilar to the replicase protein sequences of the MS2-like genogroup I strains, displaying a similarity range of 84-85%. In contrast, a similarity of 97-99% was observed among the highly conserved replicases of the nine MS2-like strains. Strain fr shared an 80% replicase similarity to JS strains and approximately 88-90% similarity to MS2-like strains

(Table 6.3). Group II replicase was approximately 50-53% similar to JS strains and to the other genogroup I strains (Table 6.4).

Table 6.1 JS strains and genogroup I. Open Reading Frame positions and genome lengths of FRNA coliphage JS strains DL52 and DL54 compared to genogroup I strains.

		Open Reading Frame Location (nt)				
Strain	**Group**	**Full length nt**	**ORF1**	**ORF2**	**ORF3**	**ORF4**
DL1	**I**	**3570**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
DL2	**I**	**3491[b]**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
DL13	**I**	**3491[b]**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
DL16	**I**	**3569**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
J20	**I**	**3569**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
ST4	**I**	**3569**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
R17	**I**	**3569**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
MS2[a]	**I**	**3569**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398**
DL52	**JS**	**3525**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398[c]**
DL54	**JS**	**3398[b]**	**130-1311**	**1335-1727**	**1678-1905**	**1761-3398[c]**

[a] GenBank prototype strain

[b] nearly full-length

[c] deletions and insertions in JS ORF4

nt - nucleotide

Table 6.2 Pairwise comparison JS strains and genogroup I. Pairwise comparisons of nucleotide percent similarity.

(A) JS strains and group I (B) JS strains and group II.

(A) Group I and JS strains.

	DL52	DL54
Strain		
DL52	100	
DL54	96.73	100
DL1	81.48	81.87
DL16	85.41	84.72
ST4	80.30	80.11
R17	80.55	80.53
J20	82.00	82.01
MS2	80.12	80.01
fr	69.18	69.06

(B) Group II and JS strains.

	DL52	DL54
Strain		
DL52	100	
DL54	96.73	100
T72	53.96	53.53
DL10	54.07	53.89
DL20	52.87	52.65
GA	52.44	52.29
KU1	52.94	52.66

The replicase protein of all genogroup I strains including the JS subgroup was shown to be 545 amino acids in length (Appendix C). The JS replicase protein was, however, distinctive from the MS2-like replicase protein as it possessed a single amino acid insertion at position 467 and aone amino acid deletion at the 3' termini of the stop codon (Appendix C). The catalytic domain of the replicase protein was in the same location, between amino acid positions 243-373. In the JS replicase proteins from amino acid position 455 and continuing to the 3' end, the sequences of the JS replicase protein diverge from the parental MS2-like strains and were unique in composition (Fig 6.3, Appendix C) resulting from a frame shift having a two nucleotide insertion (Fig 6.5).

A nucleotide alignment revealed numerous deletions in the JS strains when compared to the other genogroup I strains. Beginning approx 40 nt downstream of the replicase ORF4 stop codon and continuing to the 3' termini, 53 nt deletions were observed in the JS strains. JS strains, however, share the 3' "signature", ACCACCCA, sequence with groups I and II *Levivirus* (Appendix C).

A cluster analysis of the amino acids sequences from the replicase proteins of group I, JS subgroup and group II strains was performed. Cophenetic correlations showed the MS2-like strains including strain fr, the JS subgroup strains, and the group II strains all formed faithful clusters with correlations of 100, 90 and 98, respectively, The cluster cutoff method, however, showed only two relevant clusters, genogroup I strains, which included fr and JS, and genogroup II strains (Fig 6.4).

Table 6.3 Pairwise comparison of amino acids for JS and genogroup I. Percent similarity in amino acid

sequences between JS strains and genogroup I maturation, capsid, lysis and replicase proteins.

Maturation protein

ST4 -I	100.00											
MS2 -I	99.85	100.00										
R17 -I	98.18	98.33	100.00									
M12 -I	97.73	97.89	97.36	100.00								
DL1 -I	97.68	97.83	97.23	97.21	100.00							
J20 -I	97.85	98.01	97.11	97.07	98.26	100.00						
DL13 -I	97.45	97.61	97.13	97.45	98.07	98.11	100.00					
DL16 -I	97.45	97.61	97.13	97.45	98.07	98.11	100.00	100.00				
DL52 JS	97.45	97.61	97.13	97.45	98.07	98.11	100.00	100.00	100.00			
DL54 JS	97.45	97.61	97.13	97.45	98.07	98.11	100.00	100.00	100.00	100.00		
DL2 -I	97.63	97.79	97.31	97.63	98.25	98.11	99.83	99.83	99.83	99.83	100.00	
fr -I	84.36	84.35	84.26	84.16	84.99	84.71	84.13	84.13	84.13	84.13	84.35	100.00

Capsid protein

ST4 -I	100.00											
R17 -I	100.00	100.00										
MS2 -I	100.00	100.00	100.00									
M12 -I	98.70	98.70	98.70	100.00								
DL1 -I	98.48	98.48	98.48	98.23	100.00							
DL2 -I	98.48	98.48	98.48	98.23	100.00	100.00						
DL13 -I	98.48	98.48	98.48	98.23	100.00	100.00	100.00					
DL16 -I	98.48	98.48	98.48	98.23	100.00	100.00	100.00	100.00				
J20 -I	98.48	98.48	98.48	98.23	100.00	100.00	100.00	100.00	100.00			
DL 52 JS	98.48	98.48	98.48	98.23	100.00	100.00	100.00	100.00	100.00	100.00		
DL54 JS	98.48	98.48	98.48	98.23	100.00	100.00	100.00	100.00	100.00	100.00	100.00	
fr -I	91.20	91.20	91.20	90.92	91.85	91.85	91.85	91.85	91.85	91.85	91.85	100.00

Lysis protein

ST4 -I	100.00											
MS2 -I	100.00	100.00										
R17 -I	98.87	98.87	100.00									
DL13 -I	94.99	94.99	94.99	100.00								
DL16 -I	94.99	94.99	94.99	100.00	100.00							
DL2 -I	94.99	94.99	94.99	100.00	100.00	100.00						
DL52 JS	94.99	94.99	94.99	100.00	100.00	100.00	100.00					
DL54 JS	94.99	94.99	94.99	100.00	100.00	100.00	100.00	100.00				
DL1 -I	95.89	95.89	95.89	99.15	99.15	99.15	99.15	99.15	100.00			
J20 -I	91.13	91.13	91.13	96.40	96.40	96.40	96.40	96.40	95.49	100.00		
M12 -I	93.83	93.83	94.07	90.34	90.34	90.34	90.34	90.34	91.31	86.23	100.00	
fr -I	78.19	78.19	77.96	78.86	78.86	78.86	78.86	78.86	77.66	77.66	75.73	100.00

Replicase protein

ST4 -I	100.00											
R17 -I	99.06	100.00										
MS2 -I	99.03	98.57	100.00									
DL13 -I	97.39	97.01	97.41	100.00								
DL2 -I	97.51	97.13	97.53	99.89	100.00							
DL16 -I	97.22	96.84	97.25	99.61	99.73	100.00						
DL1 -I	97.99	97.61	97.79	99.29	99.41	99.13	100.00					
J20 -I	97.67	97.20	97.47	98.87	98.99	98.71	99.13	100.00				
fr -I	88.51	88.38	88.17	88.90	89.04	88.71	88.89	88.81	100.00			
DL52 JS	84.40	83.86	83.95	85.38	85.41	85.18	85.33	84.81	79.70	100.00		
DL54 JS	84.15	84.12	83.93	85.02	85.05	84.82	85.03	84.81	79.81	97.20	100.00	

Table 6.4 Pairwise comparison of replicase protein JS and genogroup I. Percent similarity in amino

sequences between JS strains and genogroups I and II replicase protein. Pairwise alignments were performed in Bionumerics.

Group I

ST4	100.00
R17	99.06 100.00
MS2	99.03 98.57 100.00
DL2	97.51 97.13 97.53 100.00
DL13	97.39 97.01 97.41 99.89 100.00
DL16	97.22 96.84 97.25 99.73 99.61 100.00
DL1	97.99 97.61 97.79 99.41 99.29 99.13 100.00
J20	97.67 97.20 97.47 98.99 98.87 98.71 99.13 100.00
fr	88.51 88.38 88.17 89.04 88.90 88.71 88.89 88.81 100.00

Group JS

DL52 JS	84.40 83.86 83.95 85.41 85.38 85.18 85.33 84.81 79.70 100.00
DL54 JS	84.15 84.12 83.93 85.05 85.02 84.82 85.03 84.81 79.81 97.20 100.00

Group II

T72	51.90 51.54 51.09 52.61 52.61 52.29 52.64 52.69 52.75 52.85 53.11 100.00
KU 1	51.16 50.80 50.35 51.90 51.91 51.57 51.93 51.97 52.49 52.54 52.59 97.13 100.00
DL10	52.67 52.31 51.87 53.37 53.37 53.05 53.39 53.66 52.74 53.68 53.83 93.29 93.02 100.00
DL20	52.70 52.34 51.90 53.43 53.43 53.11 53.22 53.71 53.22 53.35 53.50 93.77 93.41 98.62 100.00
GA	51.27 50.90 50.45 52.01 52.01 51.68 51.80 52.30 51.80 52.47 52.62 92.43 92.27 97.73 98.57 100.00

Figure 6.1 SimPlot analysis of genome nucleotide profile strain MS2 and DL52.

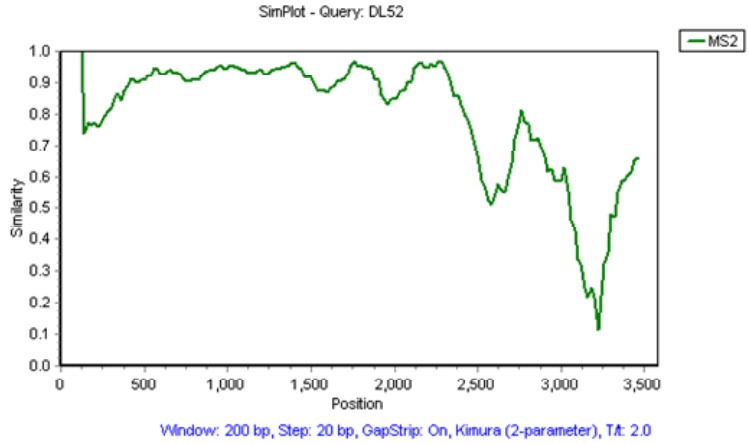


Figure 6.2 SimPlot analysis of genome nucleotide profile strain MS2 and fr.

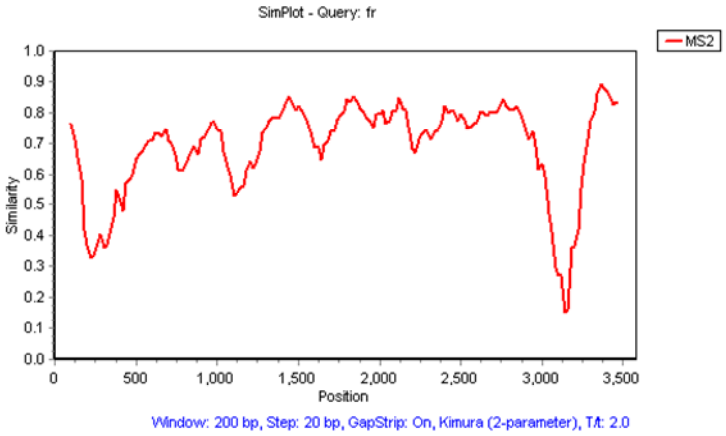


Figure 6.3 SimPlot analysis of replicase amino acid profile strain MS2 and DL52.

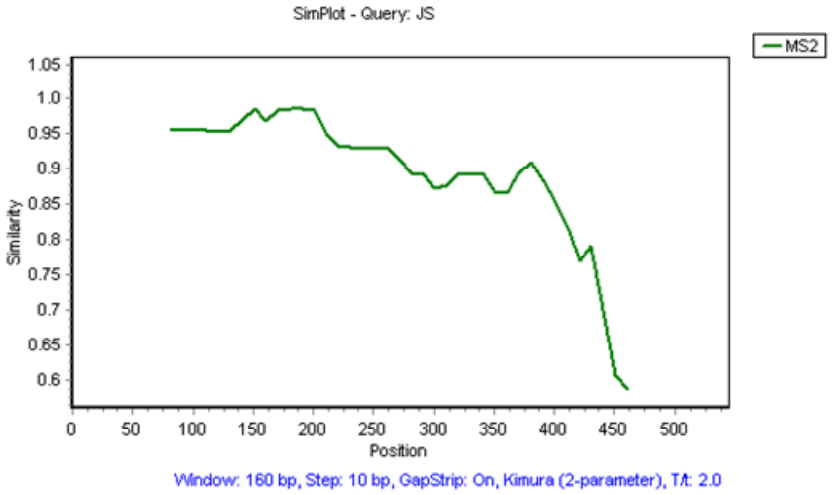


Figure 6.4 Phylogenetic analysis of replicase amino acids groups I, II and JS.

Cluster analysis of Levivirus Groups I and II phages generated from pairwise similarities of the amino acid sequences of their replicase genes. Horizontal bars at three of the branches show the standard deviations of the average similarites of the clusters. Numbers at each branch are the cophenetic correlations which represent the faithfulness of the clusters. Two relevant clusters, as determined by the cluster Cutoff method, are grouped to the right of the dashed lines.

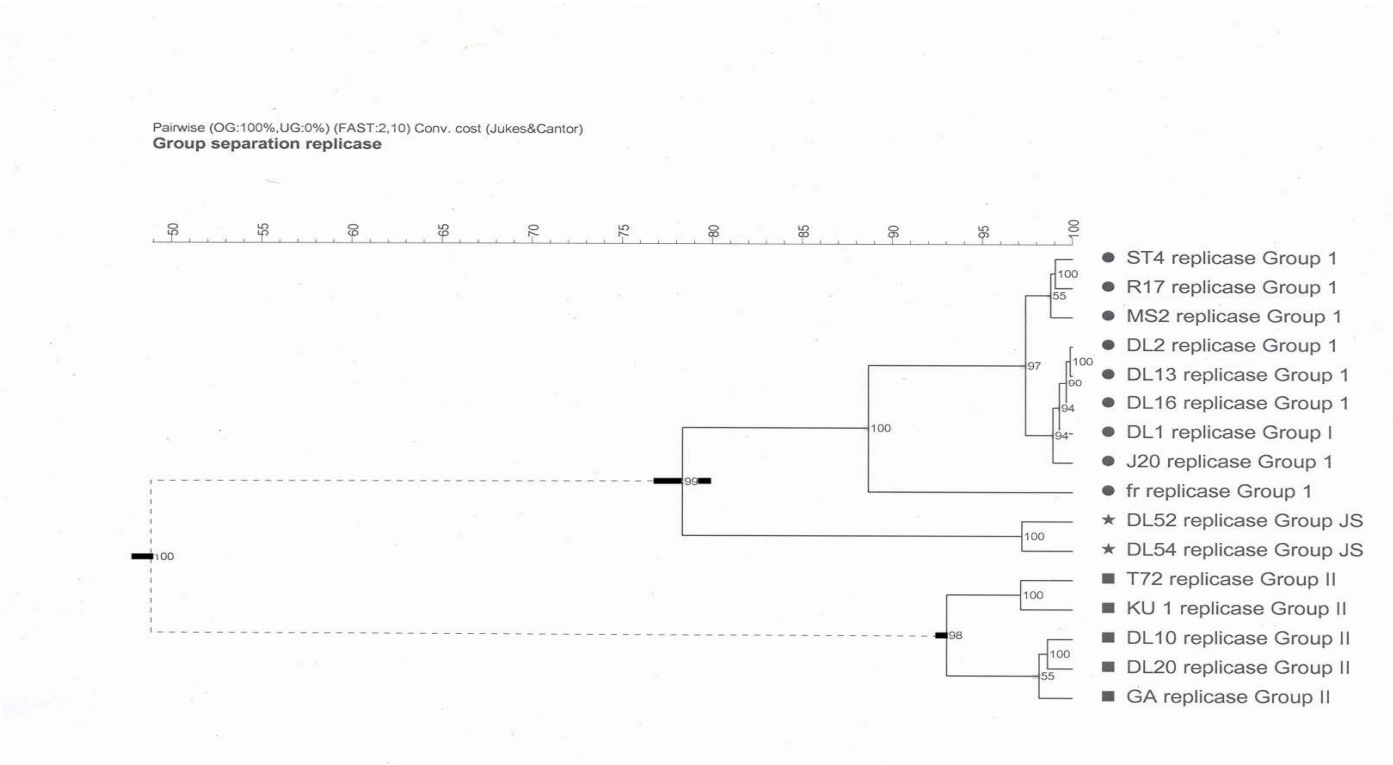


Figure 6.5 JS strains replicase frame-shift.

Replicase frame shift in two JS strains when compared to genogroup I strains. Alignment (BioEdit v7.0.1) of the replicase nucleotide sequences from genogroup I strains DL1, DL2, DL13, DL16, ST4, J20, MS2 with JS strains DL52 and DL54. For clarity, only a portion of the alignment is shown. Alignment of each genogroup is depicted in discontinuous blocks to illustrate the nucleotide position. The numbers along the top are the nucleotide positions. Genome sequences read 5' - 3' direction. Dots indicate identity with the consensus sequence. Degenerate bases are noted in the standard IUB codes. The replicase start codon and two nucleotide insertions are highlighted in red.

Discussion

The JS strains and fr diverged from the MS2-like reference strains but in different ways. Across the entire genome strain fr consistently differed from the MS2-like strains (Fig 6.2). However, with JS strains, major differences were only observed downstream of the catalytic site in the 3' end of the replicase gene and the adjacent noncoding region, suggesting a specific genetic rearrangement or recombination event (Fig 6.3).

Cophenetic correlations strengthen the possibility that JS strains are recombinants as the JS is only a subgroup of genogroup I and not a novel genogroup. Throughout the *Leviviridae* family, subgroups emerge within genogroups, however, as with strain fr, subgroup strains differ in all four genes from the parent genogroup (Fig 5.4A, B).

Genetic exchange in ssRNA viruses was first demonstrated in polioviruses (Hirst, 1962; Ledinko, 1963). Subsequent experiments with ssRNA coliphage mutants failed to provide evidence for recombination. Horiuchi (1975) concluded that RNA phages would not undergo recombination. Those attempts to detect recombination occurred in the time when FRNA phages were thought to possess only three genes, not four. In all likelihood, laboratory-applied selective pressure failed to detect or generate a specific recombinant and may not necessarily reflect the lack of recombination or responsible mechanisms that could occur under conditions that better represent the natural history and ecology of these ssRNA coliphages. Eventually, ssRNA recombination was demonstrated in a *Leviviridae* coliphage Qβ replicase (Munishkin et al., 1988).

Recombination events sometimes alter the RNA polymerase region. Human Noroviruses, a positive sense ssRNA virus with a genome length of 7400-8300 nt, are

considered to belong to a prototype strain if they share approximately 85% overall nucleotide sequence identity and a high amino acid sequence identity (> 95%) to the RNA polymerase gene (Jiang et al., 1999). The naturally occurring human Norovirus strain shared 95% amino acid sequence identity with the capsid sequences from a Mexico cluster and 95% animo acid identity to the RNA polymerase in a Lordsdale virus cluster. Sequences from the natural strain were obtained from one viral isolate. The combination of sequences in the one strain being complementary to two distinct human Norovirus clusters led to the proposition that this strain was a naturally occurring recombinant (Jiang et al., 1999).

Genetic recombination is known to occur in certain Enteroviruses, a positive ssRNA virus having an approximate 7500 nt genome. Poliovirus recombination occurs in vaccine-derived strains (Kew et al., 2002) in the human population as a single infected individual excretes a high proportion of recombinants (Oprisan et al., 2002). To determine if other enteroviruses undergo natural recombination, isolates of echoviruses were collected from a meningitis outbreak. Nucleotide sequences were clustered based on a capsid protein (VP1) and RNA polymerase (3D). Dendogram relatedness of the echovirus strains grouped the VPI sequences to the prototype strains. However, the RNA polymerase sequences did not cluster to the prototype strains, suggesting genetic recombination among the outbreak strains (Oprisan et al., 2002).

Human astroviruses are positive sense, ssRNA with a genome length of approximately 6,800 nucleotides (Walter et al., 2001) and a polyadenylated 3' tail (Belliot et al., 1997). Two sets of strains were investigated for recombination; one set was identified from a child care center in Houston, TX and the two other strains originated in stool samples from two children

in Mexico City. The pool of strains shared >97% nucleotide sequence similarity in two out of three genomic regions. The novel strain clustered to one group based on the capsid region. When the RNA polymerase gene was analyzed, the novel strain clustered to a separate human astrovirus group. The strains were identified as naturally occurring recombinants on the evidence of high sequence similarity to a few genes of one prototype and similarity to different genes in a second prototype. A total of 64 additional human astroviruses lacked these novel traits (Walter et al., 2001).

Evidence for recombination among positive ssRNA viruses exists within the RNA-dependent RNA polymerase. Turkey astrovirus is a non-enveloped, positive sense ssRNA virus with a polyadenylated 3' tailed genome of approximately 7kb. Astroviruses are associated with enteric disease (Pantin-Jackwood et al., 2006). The most conserved gene in the avian and mammalian astrovirus is the RNA-dependent RNA polymerase. Genetic analysis of capsid and polymerase sequences from twenty-three turkey astrovirus strains resulted in 8 clusters for the capsid gene and two phylogenetic clusters for the RNA polymerase gene. Computer-generated analyses identified polymerase gene recombination in strains of turkey astrovirus (Pantin-Jackwood et al., 2006).

Numerous reports of positive sense ssRNA viral recombinants are documented in the scientific literature (Cristina and Colina, 2006; Pantin-Jackwood et al., 2006; Holmes et al., 1999; Oberste et al., 2004; Banner and Lai, 1991; Oprisan et al., 2002; Walter et al., 2001; Jiang et al., 1999; Belliot et al., 1997). Virus strains are classified as natural recombinants when one virus strain is complementary to two different proteins or stretch of nucleotide sequences originating from two genetically distinct clusters or in other rearrangements events

(Chetverin, 1999). In this study, two JS strains shared >95% amino acid identity in three (maturation, capsid and lysis) *Levivirus* genes but only 84-85% amino acid identity to the otherwise highly conserved replicase protein. A nucleotide frame shift occurred downstream of the catalytic site in the replicase gene thereby accounting for the lack of nucleotide or amino acid similarity between the JS strains and the genogroup I replicase. Therefore, it is plausible to propose natural recombination in these two FRNA coliphages.

Conclusion

Phylogenetic tree analysis produced a cophenetic correlation which showed 1) ten genogroup I strains, including strain fr, 2) the JS subgroup strains, and 3) the genogroup II strains all formed faithful clusters with correlations of 100, 90 and 98, respectively. The cluster cutoff method, however, showed only two relevant clusters, 1) genogroup I strains, which included fr and JS, and 2) genogroup II strains (Fig 6.4). Therefore, the novel JS strains are not a unique *Levivirus* genogroup. The proposed classification of JS strains is a genogroup I subgroup "JS-like".

The results of this study provide molecular genetic evidence indicative of recombination in two JS strains of FRNA coliphages. There was high nucleotide and amino acid identity in three genes, the maturation, capsid and lysis genes ($\geq$ 95%) but a lack of similarity in the replicase gene. A nucleotide frame shift occurred downstream of the replicase catalytic. Therefore, the catalytic site was conserved resulting in viable progeny.

Each JS strain was isolated from two different geographical states, North Carolina and Rhode Island, and the remaining group I strains were collected from other geographical locations across the USA and Germany (Table 7.1). Geographical distinctness does not play a

role in sequence variation within the *Leviviridae* family (Table 7.1, Appendix A). Recombination may explain why *Leviviridae* strains circulate as discrete sub-groups independent of geographical location.

Although the two JS strains were sequenced in the same laboratory, these strains were field-collected by different investigators and shipped to another location where they were plaque-purified and preliminarily classified. Therefore, the possibility that contamination resulted in false recombinants seems unlikely. This is the first description of possible recombinant strains from natural isolates in ssRNA *Leviviridae* bacteriophages.

Summary

- Two novel JS strains, DL52 and DL54, were initially isolated from North Carolina and Rhode Island, respectively.

- Three genes, maturation, capsid and lysis, shared > 95% amino acid similarity when JS strains were compared to nine MS2-like genogroup I strains.

- Among nine MS2-like strains, replicase amino acid similarity was highly conserved (97-99 % amino acid identity).

- In contrast, JS strains shared 85% replicase amino acid identity to group I MS2-like strains, 80% to group I fr and 53% to group II strains.

- High (> 95%) amino acid identity in three genes and a lack of similarity in the otherwise highly conserved replicase gene along with a statistical cluster cutoff analysis suggests the JS strains are recombinants.

References

Banner LR and MMC Lai. 1991. Random nature of coronavirus RNA recombination in the absence of selection pressure. Virol, 185:441-445.

Belliot G, H Laveran, and SS Monroe. 1997. Detection and genetic differentiation of human astroviruses: phylogenetic grouping varies by coding region. Arch Virol, 142:1323-1334.

Chetverin AB. 1999. The puzzle of RNA recombination. Minireview. FEBS Lett, 460:1-5.

Chetverin AB, DS Kopein, HV Chetverina, AA Demidenko and VI Ugarov. 2005. Viral RNA-directed RNA polymerases use diverse mechanisms to promote recombination between RNA molecules. J Biol Chem, 280,8748-8755.

Cristina J and R Colina. 2006. Evidence of structural genomic region recombination in Hepatitis C virus. Virol J, 3:53

EPA. 2004. Quality assurance/quality control guidance for laboratories performing PCR analyses on environmental samples. EPA 815-B-04-001. October 2004.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95-98.

Hirst GK. 1962. Genetic recombination with Newcastele disease virus, polioviruses, and influenza. Cold Spring Harb Symp Quant Biol, 27:303-309.

Holmes EC, M Worobey and A Rambaut. 1999. Phylogenetic evidence for recombination in Dengue virus. Mol Biol Evol, 16(3):405-409.

Horiuchi K. 1975. Genetic studies of RNA phages. *In: RNA Phages*. N. Zinder, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. pp:29-50.

Jiang X, C Espul, WM Zhong, H Cuello and DO Matson. 1999. Characterization of a novel human calicivirus that may be a naturally occurring recombinant. Arch Virol, 144:2377-2387.

Kew, O., V Morris-Glasgow, M Landaverde, C Burns, J Shaw, Z Garib, J André, E Blackman, CJ Freeman, J Jorba, R Sutter, G Tambini, L Venczel, C Pedreira, F Laender, H Shimizu, T Yoneyama, T Miyamura, H van Der Avoort, MS Oberste, D Kilpatrick, S Cochi, M Pallansch, C de Quadros. 2002. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. Science, 296(5566):356-359.

Lai, M.C. 1992. RNA recombination in animal and plant viruses. Microbial. Rev 56:61-79.

Ledinko, N. 1963. Genetic recombination with poliovirus type 1 studies of crosses between a normal horse serum-resistant mutant and several guanidine-resistant mutants of the same strain. Virology, 20(1):107-119.

Lole, KS., RC Bollinger, RS Paranjape, D Gadkari, SS Kulkarni, NG Novak, R Ingersoll, HW Sheppard and SC Ray. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype c-infected serconverters in India, with evidence of intersubtype recombination. J Virol, 73(1):152-160.

Love, DC., J Vinjé, SM Khalil, J Murphy, GL Lovelace and MD Sobsey. 2008. Evaluation of RT-PCR and reverse line blot hybridization for detection and genotyping F+ RNA coliphages from estuarine waters and molluscan shellfish. J Applied Microbiol, 104:1203-1212.

Lovmar, L., C. Fock, F Espinoza, F Bucardo, A-C Syvanen and K Bondeson. 2003. Microarrays for genotyping human Group A Rotavirus by multiplex capture and type-specific primer extension. J Clin Microbiol 41:5153-5158.

Munishkin AV, LA Voronin, and AB Chetverin. 1988. An *in vivo* recombinant RNA capable of autocatalytic synthesis by Qβ replicase. Nature, 333:473-475.

Munishkin AV, LA Voronin, VI Ugarov, LA Bondareva, HV Chetverina and AB Chetverin. 1991. Efficient templates for QB replicase are formed by recombination from heterologous sequence. J Mol Biol 221:463-472.

Oberste MS, K Maher and MA Pallansch. 2004. Evidence for frequent recombination within species *Human Enterovirus B* based on complete genomic sequences of all thirty-seven serotypes. J Virol, 78(2):855-867.

Oprisan G, M Combiescu, S Guillot, V Caro, A Combiescu, F Delpeyroux and R Crainic. 2002. Natural genetic recombination between co-circulating heterotypic enteroviruses. J Gen Virol, 83:2193-2200.

Pantin-Jackwood MJ, E Spackman and PR Woolcock. 2006. Phylogenetic analysis of turkey astroviruses reveals evidence of recombination. Virus Genes, 32:187-192.

Sobsey MD, DC Love and GL Lovelace. 2006. F+RNA coliphages as source tracking viral indicators of fecal pollution. A final report submitted to the NOAA/UNH Cooperative Institute for Coastal and Estuarine Environmental Technology (CICEET).

Vinjé J, SJG Oudejans, JR Stewart, MD Sobsey and SC Long. 2004. Molecular detection and genotyping of male-specific coliphages by reverse transcription-PCR and reverse

line blot hybridization. Appl Environ Microbiol, 70(10):5996-6004.

Walter JE, J Briggs, ML Guerrero, DO Matson, LK Pickering, G Ruiz-Palacios, T Berke and DK Mitchell. 2001. Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children. Arch Virol, 146:2357-2367.

VII. A Reverse Transcription-PCR Assay to Distinguish the Four Genogroups of Male-Specific (F+) RNA Coliphages

Abstract

Identifying, managing and reducing exposure risks from fecal contamination in recreational, drinking, shellfishing and other waters and accurately assessing risk from exposure can best be attained if tools to distinguish between sources of pollution are available. The male-specific RNA coliphage (FRNA) genogroups exhibit some degree of source specificity at the human vs animal level. Reverse-transcription PCR (RT-PCR) can be effectively used for source identification if specific primer sets are designed to be capable of identifying all members within each genogroup. In this study methods were developed for a heat-release procedure that eliminated the need for RNA purification and an RT-PCR method employing genogroup-specific primers to genotype FRNA for potential use in source tracking fecal contamination. Each genogroup-specific primer set was designed from a minimum of 5 to a maximum of 10 strains of complete genome sequences per genogroup using a total genome database of 30 strains. The four genogroup-specific primer sets generated discrete PCR amplicon sizes from a variety of environmental phage strains. Cross-reactivity with strains from other genogroups was not observed.

Introduction

F-specific ssRNA (FRNA) coliphages, family *Leviviridae*, genera *Levivirus* and *Allolevivirus*, are recognized as promising indicators of both pathogenic enteric viruses and human sewage (Grabow, 2001). They generally meet the criteria for an effective indicator of sanitary quality of water (Gerba, 1987; Sobsey et al., 2005). Host-specificity displayed by their four sub-groups also renders FRNA coliphages capable of providing information regarding sources of fecal pollution (Griffin et al., 2000). Group II and III phages generally originate from human waste, whereas groups I and IV are associated with animal waste (Havelaar et al., 1986). While data has not been invariably conclusive as to fecal source vs phage grouping, this trend occurs in a majority of cases (Schaper et al., 2002; Brion et al., 2002; Cole et al., 2003; Stewart-Pullaro et al., 2006). Subgrouping was originally based on serological properties (Sundram et al., 2006). As serological typing was shown to sometimes yield inconclusive results (Hsu et al., 1995; Beekwilder et al., 1996), genotyping techniques to group FRNA phage isolates were developed as an alternative approach for fecal source determination (Hsu et al., 1995; Beekwilder et al., 1996; Griffin et al., 2000; Vinjé et al., 2004). Initially these techniques required the isolation of phages and subsequent RNA release and membrane hybridization. Recently, reverse transcription polymerase chain reaction (RT-PCR) (Dryden et al., 2006) and reverse quantitative (RTQPCR) techniques (Kirs and Smith, 2007; Ogorzaly and Gantzer, 2006) for genotyping were developed. In these reports primer design was based on a limited number of complete or partial nucleotide sequences available in the National Center for Biotechnology Information (NCBI) GenBank. Thus, broad primer and probe genogroup specificity may not have been achieved due to the limited nucleotide

data from the few strains available in GenBank. In this study a design of highly precise, forward and reverse genogroup-specific, RT-PCR primers were based on a total of 30 FRNA phages of several strains from all four genogroups. Data were based upon full genomic sequences of nineteen newly sequenced FRNA coliphages along with eleven full-length FRNA coliphage sequences available in GenBank.

Purpose

- Design genogroup-specific primer sets and develop a one-step Reverse-Transcription PCR to be capable of identifying all members within each FRNA genogroup.

Approach

- Designed genogroup-specific primer sets based on alignment of full-length genome nucleotide sequences to a minimum of 5 strains/genogroup to a maximum of 10 strains/genogroup.

- Designed primer sets to produce discrete PCR amplicon sizes specific to each genogroup.

- Evaluated primers for hairpin loops, primer-dimers and melting profiles for RT-PCR optimization.

- Optimized the heat-release of purified FRNA viral stocks to serve as a template for RT-PCR.

- Optimized the one-step format and cycling temperatures for RT-PCR.

- Validated the one-step RT-PCR with 25 FRNA phages and tested for cross-reactivity.

Materials and methods

Coliphage Isolates and Propagation

FRNA phages were collected from water, sewage and various animals representative of diverse geographical locations (Table 7.1). Vinjé et al., (2004) designed two degenerate primer sets, MJV82 and JV81 for *Leviviruses* and MJV82 and JV41 for *Allolevivirus*, specific to the replicase region. The combination of these primer sets was termed "generic PCR" as the primers were genus specific but not genogroup specific. Preliminary subgrouping of phages was previously determined by generic RT-PCR and reverse line-blot hybridization (Vinjé et al., 2004). Environmental and prototype phages were provided from The University of North Carolina at Chapel Hill collection. Strain R17 was purchased from Felix D'Herelle Reference Centre for Bacterial Viruses, Universite Laval, Quebec, Canada.

Table 7.1 Sources of *Leviviridae* strains.

Origin of *Leviviridae* Strains

Group I

Strain	Group	Source	Collected	Site collected
DL1	I	water	Jan 2004	Tijuana River, CA
DL2	I	water	Feb 2004	Delaware Bay, DE
DL13	I	oyster	Oct 2004	Whiskey Creek (Masonboro Is), NC
DL16	I	water	Nov 2004	Great Bay (Nannie Is), NH
J20	I	chicken litter	Sept 2000	South Carolina
ST4	I	UNK	UNK	UNK
R17	I	sewage	1962	Philadelphia, PA[a]
MS2	I	sewage	Sept 1959	Berkeley, CA[b]
M12	I	UNK	UNK	Germany[c]
fr	I	dung-hill	1963	Heidelberg, Germany[d]

Group II

T72	II	bird	June 2002	Talbert Marsh sandflats, Huntington Beach, CA
GA	II	sewage	Oct 1964	Ookayama, Japan[e]
KU1	II	sewage	1973	Kuwait[l]
DL10	II	mussel	April 2005	Tijuana River, CA
DL20	II	clam	May 2005	Naragansett Bay, RI

Group III

TW18	III	sewage	1970	Changhua, Taiwan[f]
HL4-9	III	hog lagoon	May 2000	Duplin County, NC
BR12	III	water	July 2005	New Market Creek Charleston, SC
VK	III	sewage	Oct 1963	Tokyo, Japan[e]
BZ1	III	sewage/feces	Oct-Nov 1971	Recife, Brazil[g]
QB	III	human feces	June 1961	Kyoto, Japan[e]
M11	III	UNK	UNK	Netherlands[h]
MX1	III	raw wastewater	1973	Campeche, Mexico[i,l]

Group IV

HB-P22	IV	bird	April 2002	Talbert Marsh sandflats, Huntington Beach, CA
HB-P24	IV	bird	April 2002	Talbert Marsh sandflats, Huntington Beach, CA
BR1	IV	water	Feb 2005	Guerin Creek (Charleston), SC
BR8	IV	water	June 2005	Bull Creek (Charleston), SC
NL95	IV	calves	UNK	Netherlands[h]
SP	IV	Siamang gibbon	1968	Tokyo, Japan[l]
F1	IV	infants	1969	Hachioji, Japan[k]

UNK = unknown

[a] Paranchych and Graham, 1962, [b] Dr. Alvin J. Clark, personal communication, [c] Zinder, 1965, [d] Marvin and Hoffman-Berling, 1963, [e] Watanabe et al., 1967, [f] Miyake et al., 1971, [g] Miyake et al., 1973, [h] Beekwilder et al., 1996, [i] Hirashima et al., 1983, [j] Sakurai et al., 1968, [k] Miyake et al., 1969, [m] Furuse et al., 1975.

UNK - unknown

Single-plaque coliphages were further purified and enriched using *Escherichia coli* HS(pFamp)R as host (Vinjé et al., 2004). Overnight enrichments were centrifuged (3,220 X g for 10 min) to pellet host cells and debris. A 1:1 chloroform to supernatant mixture was vortexed and centrifuged again. Approx 1-2 ml aliquots of the purified supernatant were frozen at −75 ^{o} C.

Coliphage titers were determined using a single agar layer procedure (SAL) (US EPA Method 1602, 2001). The procedure was as follows. A 1 ml of overnight *E. coli* Famp host culture was transferred into 50 ml trypticase soy broth with streptomycin and ampicillin (TSB/strep/amp) and grown 4 hr to log phase. A 150 ml volume of trypticase soy agar, TSA (4.5 g TSB and 1.2 g agar), was autoclaved and placed into a water bath (47-55 ^{o} C). A 300 ul aliquot of 500X strep/amp was added into the cooled 150 ml TSA. Serial 10-fold dilutions of the purified virus were prepared. For 10-fold stock dilution, 100 ul of stock was transferred into 900 ul phosphate buffered saline (PBS), mixed and this process repeated to a serial dilution 10^{-15}. To sterile15 ml plastic tubes, labeled -1 to -15, 1 ml of the 4 hr *E. coli*, 100 ul of the appropriate virus dilution and 10 ml TSA were added. After quick manual mixing the tube contents were poured into labeled 20mm petri plates which were then cooled, allowed to dry, inverted and incubated overnight at 37 ^{o} C. The coliphage titer was determined by counting the plates having a minum of 5 to a maximum of 20 plaques per plate. Titer concentration was reported as PFU/ml.

Coliphage RNA Isolation

Coliphage RNA was extracted from the purified viral supernatant using a QIAamp viral RNA mini kit (Qiagen, Valencia, CA) as follows. A 200 ul volume of purified virus

stock was added into the 25 ul protease tube. Into the same tube, 200 ul of carrier RNA/Buffer AL mixture was added and the combined solutions were vortex mixed, incubated 56 ^{o} C for 15 minute and clarified by a 1 minute centrifugation. To the same tube, 250 ul ethanol (EtOH) was added, vortex mixed and incubated 5 min at room temperature (RT). The lysate was applied into the QIAamp spin column and centrifuged 3-4 min at 6000 xg (8000 rpm) and the contents of the collection tube was discarded. To the column, 500 ul of Buffer AW2 was added, centrigued 1 min and the contents of the collection tube was discarded. To the column, 500 ul EtOH was added, centrifuged 1 min and collection tube was discarded. The column was transferred to a clean collection tube and centrifuged for 1 minute. The column was again transfered a new, nuclease-free tube and 50 ul RNase free water was added onto column and incubated 5 min at RT. The final centrifugation, (14,000 rpm) for 1 min and the recovered purified RNA was frozen at −20 ^{o}C.

Heat Release of Viral RNA

A direct heat-release procedure was applied to aliquots of undiluted viral supernatant. Ten ul of viral supernatant was heated in thin-walled 250 ul size PCR tubes for 5 min at 98 ^{o}C and chilled on ice for 2 min (Schwab et al., 1997; Vinjé et al., 2004). Aliquots of 5 ul were immediately placed into the RT-PCR mixture for amplification.

Reverse-Transcription PCR

A one-step single tube format, Qiagen One-Step RT-PCR kit (Qiagen, Valencia, CA), was used. The 50 ul reaction volume contained 24 ul RNase free water, 10 ul 5X Qiagen reaction buffer, 2 ul 10mM dNTP, 1 ul 10 uM forward primer, 1 ul 10 uM reverse primer and 2 ul Qiagen RT-PCR enzyme. To each 50 ul reaction volume, 5 ul of heat-released viral RNA

or 5 ul of purified RNA were used for reverse transcription. Thermal cycle (GeneAmp PCR System 9700, PE Applied Biosystems, Foster City, CA) conditions were as follows: 50 ^{o}C for 30 min, 95 ^{o}C for 15 min followed by 40 cycles of 94 ^{o}C for 1 min, 55 ^{o}C for 1 min, 72 ^{o}C for 1 min with a final extension of 72 ^{o}C for 10 min. Amplicons were separated by electrophoresis in 1.5% agarose gel
in 0.5X Tris-acetate-EDTA (TAE), stained with 20 ug/ml ethidium bromide and visualized under UV light (UVP, Upland, CA).

A genus-specific RT-PCR assay with primers MJV82 forward and *Levivirus* JV41 reverse or *Allolevivirus* JV81 reverse were utilized as a positive control (Vinjé et al., 2004). For each reaction, a no-template RT-PCR negative control was prepared.

PCR master mixes were prepared in a PCR hood (AirClean 600, AirClean Systems, Raleigh, NC) separate from template additions. To avoid contamination, PCR master mixes, amplification, electrophoresis and template and/or viral preparations were conducted in separate assigned rooms based on designated use.

Primer Design

Sequences (Table 7.2) were aligned using BioEdit v7.0.1 Clustal W application (Hall, 1999). Genogroup-specific primers (Table 7.3) targeting each of the four genogroups were designed to produce discrete amplicon sizes (Fig 7.1) for rapid genogroup-positive visualization. Details of the aligned sequences from which the primers were derived are shown in Fig 7.2. Primer set FRNA I was designed by alignment of isolates DL1, DL2, DL13, DL16, ST4, R17, J20 and GenBank strains MS2, M12 and fr. Primer set FRNA II was designed by alignment of phages T72, DL10, DL20 and GenBank strains GA and KU1.

Primer set FRNA III was designed by alignment of TW18, HL4-9, BR12, BZ1, VK and GenBank strains QB, M11 and MX1. Primer set FRNA IV was designed by alignment of HB-P22, HB-P24, BR1, BR8 and GenBank strains SP, NL95 and FI. Primers were evaluated for primer-dimers and hairpin loops by NetPrimer freeware (Premier Biosoft International, Palto Alto, CA) www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html).

Table 7.2 *Leviviridae* accession numbers. Accession numbers of *Leviviridae* male-specific ssRNA coliphages (FRNA) available and/or submitted to the GenBank/EMBL/DDBJ database.

Strain	Genogroup	Accession number
MS2	I	NC_001417.1
fr	I	NC_0011333.1
M12	I	AF195778
DL1	I	EU341815
DL16	I	EU341816
ST4	I	EU341817
R17	I	EU341818
J20	I	EU341819
GA	II	NC_001426.1
KU1	II	NC_002250.1
T72	II	EU372691
DL10	II	EU372692
DL20	II	EU372693
Qbeta	III	AY099114.1
M11	III	NC_004304.1
MX1	III	NX_001890.1
TW18	III	EU372694
HL4-9	III	EU372695
BZ1	III	EU372697
VK	III	EU372698
SP	IV	X07489.1
NL95	IV	AF059243.1
FI	IV	EF068134.1
HB-P22	IV	EU403427
HB-P24	IV	EU403428
BR1	IV	EU403429
BR8	IV	EU403430

Figure 7.1 Gel electrophoresis of FRNA phages using one-step RT-PCR.

Gel electrophoresis of heat-released FRNA coliphages following RT-PCR with genogroup-specific primers. (A, B) Genogroup I, 142 bp (C, D) Genogroup II, 471 bp (E) 1 kb Plus Track-It Ladder (F, G) Genogroup III, 795 bp (H, I) Genogroup IV, 1159 bp and (J) negative control.

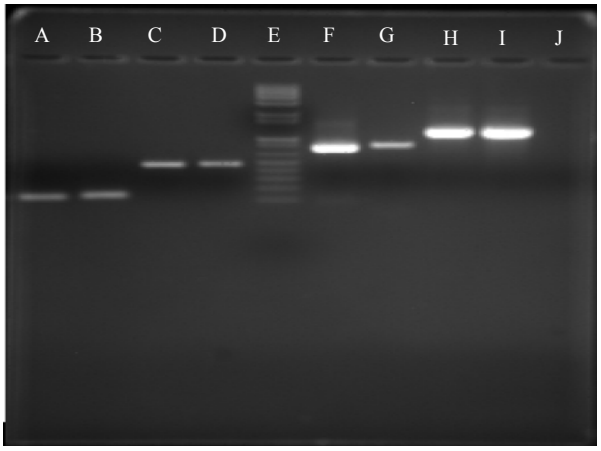
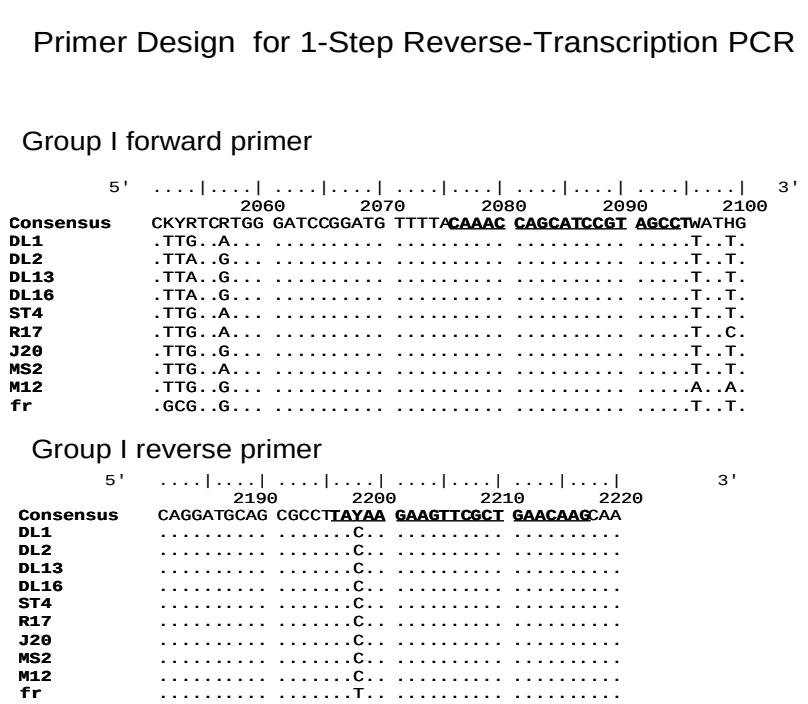


Figure7.2 Alignment and primer design of FRNA phages for one-step RT-PCR.

Primer design based on alignment of genome sequences and nucleotide positions from each respective *Leviviridae* male-specific FRNA genogroup. Alignment for each genogroup is depicted in discontinuous blocks to illustrate the nucleotide position of the primers. The numbers along the top are the nucleotide positions. Primers are underlined. Genome sequences read 5' - 3' direction. Dots indicate identity with the consensus sequence. Degenerate bases are noted in the standard IUB codes. (A) Genogroup I. (B) Genogroup II. (C) Genogroup III. (D) Genogroup IV.

(A) Group I



(B) Group II

Group II forward primer

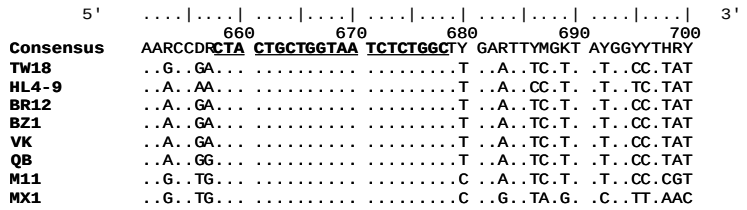
```
     5'     ....|....| ....|....|  ....|....|      3'
            90        100       110          120

Consensus   ATGCCGTTAG GTTTAGRTGA  CGGTATRTTC
T72         .......... ......G...  ......A.-.
DL10        .......... ......A...  ......A.-.
DL20        .......... ......A...  ......A...
GA          .......... ......G...  ......G.-.
KU1         .......... ......A...  ......A.-.
```

Group II reverse primer

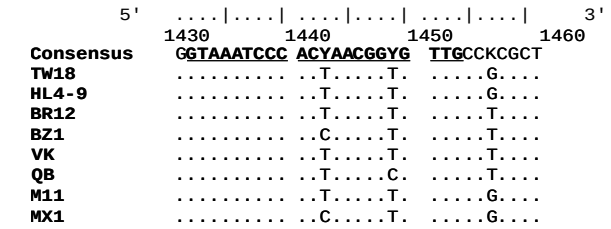
```
      5'     ....|....| ....|....| ....|....|      3'
            540       550       560       570
Consensus   TCAATATTGG GGTTGCDATT GCDGARGCYA
T72         .......... ......G... ..A..A..T.
DL10        .......... ......A... ..T..A..T.
DL20        .......... ......G... ..T..G..T.
GA          .......... ......A... ..T..A..C.
KU1         .......... ......T... ..G..A..C.
```

(C) Group III

Group III forward primer



Group III reverse primer



(D) Group IV

Group IV forward primer

```
     5'      ....|....| ....|....| ....|....| ....|....| ....|....|      3'
                   2960       2970       2980       2990       3000
Consensus    TCGTGGAAGC ATGCCTGTCC GCAGGATGTW ACCAAACGYG CRYTVAARTA
BR1          .......... .......... .........T ........C. .AT.C..G..
BR8          .......... .......... .........T ........C. .AC.C..G..
HB-P22       .......... .......... .........T ........C. .AT.C..A..
HB-P24       .......... .......... .........A ........T. .AC.G..A..
SP           .......... .......... .........T ........C. .AT.C..G..
NL95         .......... .......... .........A ........C. .AC.A..G..
FI           .......... .......... .........T ........C. .GC.G..G..
```

Group IV reverse primer

```
     5'      ....|....| ....|....| ....|....| ....|....| ....|....|       3'
                  4110       4120       4130       4140       4150
Consensus    TDGCGTGGAT TCAGGACAGT GCCTTCATCC GGCCCCCKTA TTYNMTTAMG
BR1          .G........ .......... .......... .......G.. ..TCA...C.
BR8          .A........ .......... .......... .......G.. ..TCA...C.
HB-P22       .G........ .......... .......... .......G.. ..TTC...C.
HB-P24       .G........ .......... .......... .......T.. ..TTA...A.
SP           .G........ .......... .......... .......G.. ..TAA...C.
NL95         .T........ .......... .......... .......T.. ..TTA...A.
FI           .G........ .......... .......... .......G.. ..CG---.C.
```

Results

Detection of the Four Genogroups of FRNA Coliphages

Within the aligned genomic sequences of each FRNA coliphage genogroup, areas of genetic variability were observed (Appendix A). Robust, genogroup specific RT-PCR primers were designed to conserved sequences from a variety of strains (10 strains from group I, 5 strains from group II, 8 strains from group III and 7 strains from group IV) (Table 7.1, Fig 7.2).

The four genogroup-specific primer sets generated discrete PCR amplicon sizes (Fig 7.1). The discrete amplicon sizes allowed rapid visualization of genogroup-positive FRNA strains. In addition, primer sets lacked primer-dimers or hairpin loop formations thereby allowing optimal PCR amplification. Primer set FRNA I (Table 7.3) produced 142 base pair (bp) amplicons from isolates DL1, DL2, DL13, DL16, ST4, R17, J20 and strain MS2. Primer set FRNA II produced 471 bp amplicons from isolates T72, DL10, DL20 and strain GA. Primer set III produced 795 bp amplicons from isolates TW18, HL4-9, BR12, BZ1, VK and strain QB. Primer set IV produced 1159 bp amplicons from isolates HB-P22, HB-P24, BR1, BR8 and strains SP and NL95.

Table 7.3 Genogroup-specific primer sets. Genogroup-specific primer sets designed to detect the FRNA coliphage (*Leviviridae)*. Degenerate bases are highlighted in bold and written in the standard IUB code. UTR – untranslated region.

Group	Primer	Sequence	Amplicon (bp)	Gene
I	FRNA I F (forward)	5' CAAACCAGCATCCGTAGCC 3'	141	Replicase
I	FRNA I R (reverse)	5' CTTGTTCAGCGAACTTCTT**R**TA 3'		Replicase
II	FRNA II F (forward)	5' ATGCCGTTAGGTTTAG**R**TGAC 3'	471	5' UTR
II	FRNA II R (reverse)	5' GCAAT**H**GCAACCCCAATA 3'		Maturation
III	FRNA III F (forward)	5' CTACTGCTGGTAATCTCTGGC 3'	795	Maturation
III	FRNA III R (reverse)	5'CAAC**R**CCGTT**R**TGGGATTTA C 3'		Capsid
IV	FRNA IV F (forward)	5' CTGTCCGCAGGATGT**W**ACCA 3'	1159	Replicase
IV	FRNA IV R (reverse)	5' GGCACTGTCCTGAATCCACG 3'		Replicase

Detection of FRNA Coliphages Using Heat-released Viral RNA

For most coliphage strains, heat-released viral RNA (Schwab et al., 1997; Vinjé et al., 2004) produced the desired amplicon. If the amplicon was not observed, purified RNA was used to supplement the RT-PCR assay. Strains requiring purified RNA were J20, GA, QB and NL95. Use of RNA vs heat-release was not titer dependent as phage QB had a titer > 10^{10} but this strain would not amplify with heat-release whereas strain SP amplified at a 10^{3} titer.

Testing for Cross-reactivity of Primer Sets

Although BLAST (www.ncbi.nlm.nih.gov/BLAST) results did not indicate cross-reactivity in GenBank strains, each primer set was tested using a total of 25 environmental and prototype FRNA strains representing each genogroup. Cross-reactivity was not observed between groups I and II (genus *Levivirus*) or between groups III and IV (genus *Allolevivirus*). However, primer set IV produced faint non-specific amplification when *Levivirus* genogroups I and II strains were tested with RT-PCR (data not shown). Preliminary screening of environmental FRNA strains using generic primers as described (Vinjé et al., 2004) will determine the genus. Once the genus is determined by generic RT-PCR, subsequent genogroup-specific primers, I and II for *Levivirus* or III and IV for *Allolevivirus,* will eliminate non-specific amplification. However, the non-specific PCR amplification products were faint and amplicons were not the same molecular size or intensity as generated by each genogroup-specific primer set.

Discussion

In this study are reported the first broadly representative but specific RT-PCR primer

sets for each of the four separate FRNA coliphage genogroups designed from a large (30 strains) genetic sequence database. The robust genogroup-specific primers detected a variety of purified FRNA coliphages from various sources and samples collected around the world. Diverse FRNA coliphage sources were adequately represented, as those used for primer design came from a variety of water bodies, fecal waste sources and their animal hosts including birds, chickens, swine, oyster, mussel, clam, sewage, human feces, calves and lesser apes (Table 7.1). Each genogroup-specific primer set was designed from a minimum of 5 strains per group in which the complete genome was known. In addition, FRNA genogroup strains were isolated from the United States, Japan, Europe, Brazil, Taiwan and/or Mexico.

Previous investigations using real-time PCR or RT-PCR for detection and genotyping of FRNA coliphages used extracted viral RNA from either purified virus or environmental water samples (O'Connell et al., 2006; Dryden et al., 2006; Kirs and Smith, 2007; Ogorzaly and Gantzer, 2006). In this study, coliphage RNA was made available for amplification by RT-PCR using a heat-release RNA technique directly applicable to culture enriched coliphage isolates (Schwab et al., 1997; Vinjé et al., 2004). This heat-release procedure reduces RT-PCR preparation time by omitting initial coliphage isolation followed by RNA chemical purification steps. In field samples where low numbers of coliphages are present, it may be necessary to first enrich for FRNA coliphages (Love and Sobsey, 2007) followed by heat-release of enrichments for RT-PCR amplification.

Using primers designed from only a limited number of complete or partial nucleotide sequences available in the NCBI GenBank, Dryden et al., (2006), failed to detect FRNA coliphages in several environmental samples, despite the fact that coliphages were detected at

all sites by the single agar layer method (SAL). The authors acknowledged the possibility that their RT-PCR assay failed to detect unknown FRNA phages. Indeed, this may have been the case as their primer sets were designed to the limited FRNA sequences in GenBank, and, primers were not designed to genogroup III.

In previous studies, the sensitivity of lower detection limits appears to be generally consistent. Using purified RNA, RTQPCR for all four genogroups of FRNA phages detected 0.1 plaque-forming units (PFU) per 50 ul reaction in seawater and 0.5 PFU/50 ul reaction in stool samples (Kirs and Smith, 2007). These findings were comparable to a sensitivity of 0.1 PFU in laboratory-prepared MS2 in an RT-PCR assay using a single primer set that detects, but does not differentiate, the *Leviviridae* family (Rose et al., 1997). Lower detection limits for various FRNA phage strains with RTQPCR ranged from 1-10 PFU/ml for MS2, 0.01-0.1 PFU/ml for GA and SP and 0.1-1 PFU/ml for phage QB (Ogorzaly and Gantzer, 2006). The sensitivity of a real-time fluorogenic RT-PCR for FRNA strain MS2 detected 40-0.4 fg of RNA per 20 ul reaction (O'Connell et al., 2006). Sensitivity of detection based on a lower limit or dilution endpoint to extinction was not measured in this assay. The method was applied to phage enrichments that contained $>10^3$ PFU. Further studies are planned to determine the lower limit of FRNA coliphage detection and genotyping when the method is applied directly to enriched and serially diluted coliphages and coliphages in unenriched environmental samples such as feces, manure, biosolids and wastewater.

The new genogroup-specific primers, RNA preparation and RT-PCR amplification procedures described here should facilitate improved and more reliable genotyping of different FRNA coliphage isolates in environmental samples.

Summary

- Using genogroup-specific primers directed to highly conserved consensus sequences in each genogroup, a one-step RT-PCR was developed that detected all four genogroups when applied to 25 environmental and prototype FRNA strains.

- Primer sets produced PCR amplicon sizes of 142 bp, 471 bp, 795 bp and 1159 bp to genogroups I, II, III and IV, respectively.

- Discrete amplicon sizes allow easy visualization of genogroup-specific amplicons produced in the RT-PCR.

- The RT-PCR successfully amplified 25 FRNA coliphages when applied to direct heat-released viral RNA template, which reduced RNA sample preparation time by omitting the time-consuming and costly chemical RNA purification steps.

- These genogroup-specific primers sets can aid in source-tracking FRNA coliphages.

References

Beekwilder, J., Nieuwenhuizen, R., Poot, R., van Duin, J., 1996. Secondary structure model for the first three domains of QB RNA. Control of A-protein synthesis. J. Mol. Biol. 256(1), 8-19.

Brion, G.M., Meschke, J.S., Sobsey, M.D., 2002. F-specific RNA coliphages: occurrence, types, and survival in natural waters. Water Res. 36, 2419-2425.

Cole, D., Long, S.C., Sobsey, M.D., 2003. Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. Appl. Env. Micro. 69, 6507-6514.

Dryden , S.K., Ramaswami, B., Yuan, Z., Giammar, D.E., Angenent, L.T., 2006. A rapid reverse transcription-PCR assay for F+ RNA coliphages to trace fecal pollution in Table Rock Lake on the Arkansas-Missouri border. Water Res. 40, 3719-3724.

Furuse, K., Ando, A., Watanabe, I. 1975. Isolation and grouping of RNA phages VII. A survey in Peru, Bolivia, Mexico, Kuwait, France, Australia and the United States of America. J. Keio. Med. Soc., 52, 353-361.

Gerba, C.P., 1987. Phage as indicators of fecal pollution, in: Goyal, S.M., Gerba, C.P., Bitton, G. (Eds), Phage Ecology. Wiley Interscience, New York, pp. 197-210.

Grabow, W.O.K., 2001. Bacteriophages: update on applications as models for viruses in water. Water S.A. 27, 251-268.

Griffin, D.W., Stokes, R., Rose, J.B., Paul, J.H. III., 2000. Bacterial indicator occurrence and the use of an F(+) specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida. Microbial. Ecol. 39, 56-64.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95-98.

Havelaar, A.H., Furuse, K., Hogeboom, W.M., 1986. Bacteriophages and indicator bacteria in human and animal faeces. J. Appl. Bacteriol. 60, 255-262.

Hirashima A., Inokuchi,Y., Harigai, H., Furuse, K., Watanabe, I., 1983. Biochemical characterization of RNA coliphage MX1. J. Gen. Virol. 64(4), 939-946.

Hsu, F.C., Shieh,Y.S., van Duin, J., Beekwilder, M.J., Sobsey, M.D., 1995. Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl. Env. Micro. 61(11), 3960-3966.

Kirs, M., Smith, D.C., 2007. Multiplex quantitative real-time reverse transcriptase PCR for F+

specific RNA coliphages: a method for use in microbial source tracking. Appl. Env. Micro. 73(3), 808-814.

Love, D.C., Sobsey, M.D., 2007. Simple and rapid F+ coliphage culture, latex agglutination, and typing assay to detect and source track fecal contamination. Appl. Env. Micro. 73(13), 4110-4118.

Marvin, D.A., Hoffman-Berling, H., 1963. Physical and chemical properties of two new small bacteriophages. Nature. 197(4866), 517-518.

Miyake, T., Shiba, T., Sakurai, T., Watanabe, I., 1969. Isolation and properties of two new RNA phages SP and FI.. Japan. J. Micro. 13(4), 375-382.

Miyake, T., Furuse, K., Shiba, T., Aoi, T., Sakurai, T., Watanabe, I., 1971. Isolation and grouping of RNA phages in Taiwan. J. Keio Med. Soc. 48, 25-34. (Japanese)

Miyake, T., Furuse, K., Shiba, T., Aoi, T., Sakurai, T., Watanabe.I., 1973. Isolation and grouping of RNA phages. II. A survey in Brazil. J. Keio Med. Soc. 50, 353-362. (Japanese)

O'Connell, K.P., Bucher, J.R., Anderson, P.E., Cao, C.J., Khan,A.S., Gostomaki, M.V., Valdes, J.J., 2006. Real-time fluorogenic reverse transcription-PCR assays for the detection of bacteriophage MS2. Appl. Env. Micro. 72(1), 478-483.

Ogorzaly, L., Gantzer,C., 2006. Development of real-time RT-PCR methods for specific detection of F-specific RNA bacteriophage genogroups: application to urban raw wastewater. J. Virol. Meth. 138, 131-139.

Paranchych, W., Graham, A.F., 1962. Isolation and properties of an RNA-containing bacteriophage. J. Cell Comp. Physiol. 60, 199-208.

Rose, J.B., Zhou, X., Griffin, D.W., Paul, J.H., 1997. Comparison of PCR and plaque assay for detection and enumeration of coliphage in polluted marine waters. Appl. Env. Micro., 63(11), 4564-4566.

Sakurai, T., Miyake, T., Shiba, T., I. Watanabe, 1968. Isolation of a possible fourth group of RNA phage. Japan. J. Micro. 12, 544-546.

Schaper, M., Jofre, J., Uys, M., Grabow, W.O., 2002. Distribution of genotypes of F-specific RNA bacteriophages in human and non-human sources of faecal pollution in South African and Spain. J. Appl. Micro. 92, 657-667.

Schwab, K.J., Estes, M.K., Neill, F.H., Atmar, R.L., 1997. Use of heat release and an internal RNA standard control in reverse transcription-PCR detection of Norwalk virus from

stool samples. J. Clin. Micro. 35(2), 511-514.

Sobsey, M.D., Love, D., Lovelace, G., Stewart, J.R., Robinson. B., 2005. Methods to detect and genotype coliphages in water and shellfish. Biennial Meeting of the Interstate Shellfish Sanitation Conference. http://www.unc.edu/sobseylab/ISSCcoliphagedemo.pdf

Stewart, J.R., Daugomah, J.W., Chestnut, D.E., Graves, D.A., Sobsey, M.D., Scott, G.I., 2006. FRNA coliphage typing for microbial source tracking in surface waters. J. Appl. Micro. 101, 1015-1026.

Stewart, J.R., Vinjé, J., Oudejans, S.J.G., Scott, G.I., 2006. Sequence variation among group III F-specific RNA coliphages from water samples and swine lagoons. Appl. Env. Micro. 72, 1226-1230.

Sundram, A., Jumanial, N., Ehlers, M.M., 2006. Genotyping of F-RNA coliphages isolated from wastewater and river samples. Water S.A. 32(1), 65-70.

US EPA., 2001. Method 1602: Male-specific (F+) and somatic coliphage in water by single agar layer (SAL) procedure. Office of Water, EPA 821-R-01-029. Washington, D.C., 20460. www.epa.gov/nerlcwww/1602ap01.pdf

van Duin, J., 1998. The single-stranded RNA bacteriophages. in: Fraenkel Conrat, H., Wagner, R.R. (Eds). The Bacteriophages. Series *The Viruses*. Plenum Press, NY. pp. 117-167.

Vinjé, J., Oudejans, S.J.G., Stewart, J.R., Sobsey, M.D., Long, S.C., 2004. Molecular detection and genotyping of male-specific coliphages by reverse-transcription-PCR and reverse line blot hybridization. Appl. Env. Micro. 70(10), 5996-6004.

Watanabe, I., Miyake, T., Sakurai, T., Shiba, T., Ohno.T., 1967. Isolation and grouping of RNA phages. Proc. Jap. Acad. 43(3), 204-209.

Zinder, N.D., 1965. RNA phages. Ann. Rev. Microbiol. 19, 455-473.

VIII. Overall Discussion

Male-specific RNA coliphages, among the smallest autonomous viruses known, are positive sense, single-strand RNA (ssRNA) phages possessing a genome, 3.8 to 4.2 kb, enclosed within a non-enveloped 26 nm icosahedral-shaped capsid (Buchen-Osmond, 2003). These coliphages belonging to the family *Leviviridae* were initially grouped primarily according to their serological properites. The *Leviviridae* family contain two genera, *Levivirus* and *Allolevivirus* which are further subdivided into four major serogroups, I, II, III, IV and branched subgroups (a,b,c,d) according to serological cross-reactivity (Sundram et al., 2006). *Levivirus* are subdivided into genogroups I and II and *Allolevivirus* are subdivided into genogroups III and IV. It became apparent as early as the 1970s that the four genogroups of FRNA coliphages had somewhat different fecal source and geographic distributions. In the 1990s the development of genotyping methods based on synthetic oligonucleotide probes (Hsu et al., 1995; Beekwilder et al., 1996) made it possible and convenient to genotype FRNA coliphages and better understand their ecology, their value as fecal and viral indicators, their ability to distinguish human from animal fecal wastes and the impacts of these different waste sources on ambient waters. By employing these methods it became apparent that more information regarding source of fecal pollution could be obtained by comparing full-length genomic sequences from FRNA coliphages collected from various animals and water bodies.

Male-specific coliphages have been suggested as a viral indicator for: (1) fecal contamination (Osawa, 1981; Furuse, 1983), (2) enteric bacterial contamination (Gerba, 1987), (3) enteric viral contamination (Grabow, 2001; Leclerc et al, 2000) and (4) risks of gastro-intestinal illness from recreational water exposures (Colford et al., 2007). FRNA coliphages are almost indistinguishable from most human enteric viruses (Grabow, 2001), occur in higher numbers in sewage and wastewater effluents than viral enteric pathogens (Grabow, 2001), their presence implies the presence of pathogenic viruses (Grabow, 2001), and, in a majority of cases, they display fecal-source specificity (Vinjé et al, 2004; Cole et al, 2003; Furuse, 1987; Schaper et al, 2002; Scott et al, 2002; Stewart, 2002; Long et al, 2005).

The focus of this study was to develop and validate a rapid, genogroup-specific molecular assay for the detection of FRNA coliphages as a potential viral indicator of fecal pollution. Before the molecular assay could be developed, a genetic sequence database was generated representing environmental and prototype FRNA coliphage strains from all four genogroups.

A ssRNA viral assay would need a large (at least 5 strains/genogroup) genetic sequence database. To develop a genetic database, 19 FRNA strains were sequenced and compared to the 11 FRNA full-length sequences available in the National Center for Biotechnology Information (NCBI) genetic database (GenBank) for a total of 30 FRNA strains. FRNA phages were collected from water, sewage, and various animals representative of diverse geographical locations (Table 7.1). The field-collected FRNA strains and prototype strains were represented by phages isolated from the United States, Japan, Europe, Brazil, Taiwan and/or Mexico (Table 7.1). The majority of groups II (80%) and III (75%)

strains were collected at municipal sewage sources or water bodies with the exception of one group II strain (collected from bird droppings), one group III strain (collected from swine lagoon) and one group III strain was from an unknown source. Two out of 10 group I strains (20%) were collected from sewage, 4 out of 10 group I strains (40%) were collected from ambient waters and/or sentinel organisms (oyster, clam, mussel), one strains' source was unknown and one strain was collected from a dung-hill. Four out of seven group IV strains (57%) were collected from animal sources (bird, gibbon, calves), two strains (29%) were collected from ambient water sources and one strain was obtained from an infant (Table 7.1).

Phages were sequenced by primer walking. A polyadenylated (Poly-A) tail was added to the 3' end of purified viral RNA, the poly-A RNA was reverse transcribed with a Poly-T reverse primer; the resulting cDNA was used as a PCR template. A gene-specific forward and a poly-T primer used in the PCR mixture produced an approximate one kilobase (kb) amplicon. PCR products were gel-purified (GenScript, Piscataway, NJ), cloned (Invitrogen, Carlsbad, CA) and sequenced (Sequetech, Mountain View, CA). This process was repeated until the majority (all sequences except 100-200 bp at the 5' end) of the genome was obtained. A rapid amplification of cDNA ends (RACE) Smart RACE kit (Clontech, Mountain View, CA) was used to amplify the 5' portion of the genome.

When full-length genome nucleotide sequences were aligned with the published GenBank strains within their respective genogroup, very similar or identical gene mapping, or Open Reading Frames (ORF), were observed for all four *Leviviridae* genes indicating the sequence data generated in this study was valid. Sequence similarity among genogroup I strains ranged from 75.27 - 96.67 % with strain fr forming a separate subgroup (Table 5.2, Fig

5.2). Among group II strains, nucleotide sequence similarity ranged from 83.30 to 93.84% with strains DL10, DL20 and GA having the highest sequence identities (93.43-93.67%) whereas strains T72 and KU1 formed a separate subcluster (Table 5.2, Fig 5.2). Among *Allolevivirus* group III, two different subclusters were formed. The first subcluster was composed of strains VK, HL4-9, BR12, BZ1, TW18 and GenBank strain Qβ having a nucleotide sequence similarity ranging from 91.87-95.69%. The second subcluster formed with GenBank group III strains MX1 and M11 having an 87% nucleotide similarity to each other. The nucleotide similarity of strains between the two group III subclusters ranged from 69.77-71.33% (Table 5.2, Fig 5.2). Group IV *Allolevivirus* shared sequence identities ranging from 74.90-95.03 % with the closest identities being 95.03% between strains BR8 and BR1.

Amino acid composition was similar among genogroups, further validating the nucleotide sequences. With the exception of strain fr, amino acid number was consistent in each of the four protein types in group I phages (Table 5.4, Appendix B). The capsid protein of all strains in group I was 130 amino acids in length. *Levivirus* groups I and II capsid proteins shared a conserved region consisting of a 10 amino acid, FVLVDNGGTG, consensus sequence. Groups I and II maturation protein shared a consensus region RWLELQ at amino acid (aa) positions number 198-203. The length of the maturation protein of groups III and IV varied from 420 aa to 450 aa (Table 5.4) and possessed a mutual conserved aa region LWLEFRYGL (Appendix B). The length of the capsid protein was 133 and 132 aa for groups III and IV, respectively, and conserved stretches of amino acids occurred in both groups.

An algorithmic approach was selected to construct phylogenetic trees from the nucleotide sequence and amino acid data (Fig 5.2, Fig 5.4). Nucleotide sequences in the

phylogenetic tree of *Levivirus* group I strains produced two branches, with 9 strains clustered as MS2-like and strain fr an individual branch (Fig 5.2). Within group II nucleotide sequences, strains KU1 and T72 formed one branch and strains DL10, DL20 and GA formed a second branch. *Allolevivirus* group III nucleotide sequences clustered into a MX1, M11 branch and a second branch with Qβ-like strains BR12, VK, BZ1, HL4-9, TW18 and prototype Qβ. Nucleotide sequence analysis formed three branches in group IV strains as follows: 1) HB-P24, HB-P22 and prototype NL95, 2) BR1, BR8 and prototype SP and 3) prototype FI (Fig 5.2). Individual proteins were clustered into phylogenetic trees. In some cases, phylogenetic protein trees formed more subclusters or branches that the nucleotide trees (Fig 5.4). Genome organization, amino acid conservation and identical or very similar nucleotide start and stop positions supported the *Leviviridae* genogroup designation. In addition, eight nucleotides on the 3' termini clearly distinguish the *Allolevivirus*, 5' TCCTCCCA 3', from the *Levivirus*, 5' ACCACCCA 3'.

In addition, two new undescribed *Levivirus* strains which did not hybridize to previously designed geno-specific hybridization probes (Vinjé et al., 2004) were sequenced. The two unique FRNA strains were collected from North Carolina and Rhode Island. Full-length genomic sequences from the novel strains were compared to nucleotide and/or amino acid sequences from 10 group I strains (MS2, DL1, DL2, DL13, DL16, ST4, R17, J20, M12, fr) and 5 group II strains (T72, DL10, DL20, GA, KUI). Based on full-length genome sequences and phylogenetic analyses, these novel strains were placed into a "JS" subcluster of genogroup I. Sequence similarities of the maturation, capsid and lysis proteins of the JS strains were very similar to those of the MS2-like group I strains, sharing 99-100%, 98-100%

and 95-100% sequence similarities, respectively (Table 6.3, Fig 6.1). However, the replicase protein sequences of the JS strains were quite dissimilar to the replicase protein sequences of the MS2-like genogroup I strains, displaying a similarity range of 84-85% and a frame shift resulting from a two nucleotide insertion (Fig 6.5). The resuls of this study provide molecular genetic evidence indicative of recombination in two JS strains of FRNA coliphages. The JS strains provided insight to phage ecology and recombination events in natural FRNA strains.

In this study, analyses of complete genomic sequences from 30 FRNA phages plus two novel strains support the known classification scheme. That is, the *Leviviridae* consist of two genera and four distinct genogroups. From this analysis an observation was made that to better define the sub-groupings, it may be more reasonable to assign an association to a specific strain name, i.e., Qβ-like instead of genogroup III, subgroup a; MX1-like instead of genogroup III, subgroup b; and in group IV SP-like instead of genogroup IV, subgroup a; and FI-like instead of genogroup IV, subgroup b. Thus, a recommendation based on these findings would be to dismiss the alphabetical sub-grouping nomenclature.

Rose et al (1997) designed a one-step reverse transcription polymerase chain reaction (RT-PCR) using a single primer set that detects, but does not differentiate, FRNA coliphages. In this dissertation, a RT-PCR was designed to distinguish the four FRNA coliphage groups (I, II, III, IV) and ultimately, to distinguish human vs animal fecal sources. Primer sets were designed based on the complete genomic sequences of 30 FRNA strains. Genogroup specific RT-PCR primers were designed to conserved sequences from a variety of strains (10 strains from group I, 5 strains from group II, 8 strains from group III and 7 strains from group IV) (Table 7.1, Fig 7.2). Unique amplicon sizes were generated to allow rapid visualization of

each genogroup (Fig 7.1). The traditional one-step RT-PCR was developed, optimized for use with heat-released viral nucleic acid and tested for cross-reactivity. Rigorous validation to ensure lack of cross-reactivity was performed whereby each primer set was tested against 25 environmental and prototype strains. This assay was developed, in part, to allow laboratories lacking real-time equipment the ability to genotype FRNA isolates.

A limitation of molecular detection is that nucleic acid presence or persistence of the phage is detected and is not necessarily representative of the presence of infectious viruses. It has been suggested that if the virion capsid is disrupted, the RNA should degrade rapidly under environmental conditions. This position has not been supported by some lab and field studies which addressed long-term persistence of viral nucleic acids in environmental waters (Kirs and Smith, 2007). Another proposed approach is the discrimination between an intact but non-infectious FRNA phage based on degradation of accessible viral RNA by RNase to eliminate detection of free RNA from damaged (leaky) capsids.

The molecular FRNA phage assay developed in this study may be applicable to an accelerated turnaround time as the assay omits the RNA purification procedure by use of direct heat-release. The traditional primers developed not only allow genogroup identification but provide a comprehensive assessment as to the sanitary quality of the water. If any FRNA phages are detected, then fecal contamination has occurred. This approach utilized a non-cultivation library-independent method for differentiating between human and animal fecal contamination.

In a global phage genotyping assessment, genotypes were reported to be differentially distributed. For example, the FRNA phage from sewage samples in Brazil and West

Germany were from group I exclusively. However, it was unclear as to whether or not sewage treatment plants received slaughterhouse waste (Furuse, 1987). If, however, both slaughterhouse and domestic sewage were combined, and, if only groups II and III are specific to humans, then presumably at least genogroups I, II and III would have been detected in their study. Furuse argued that group I phages observed in raw sewage from treatment plants were most likely introduced from animal sources, and group II and group III phage were from human sources. This begs the question as to why their study only detected group I in sewage treatment facilities from Germany and Brazil. However, several explanations are possible for this result. Group I could have out-competed groups II and III or they could have slower inactivation rates. This trend of persistent goup I FRNA coliphages should be apparent in other sampling stations if these explanations are correct. However, only limited genotyping from other studies are available for making such observations (Osawa et al., 1981; Miyake et al., 1971). Certain FRNA genogroups may predominate in various human and/or animal populations and their occurrence may also be influenced by climate, diet, intestinal fauna, etc. Further study is needed in different geographical locations to better understand the ecology or natural history of the different FRNA coliphage genogroups; methods developed in the current study should contribute to this effort. Despite the lack of an absolute association between an FRNA genogroup and a unique source, these coliphages likely signal the presence of fecal contamination from either animal and/or human origin. Thus, an FRNA positive sample(s) warrants further investigation, intervention and under some circumstances a public notification alert.

The following three paragraphs will discuss governmental and organizational standards for sanitary quality of recreational water as they pertain to applications of this dissertation. Many of our nation's ambient water resources are impaired and fail to achieve US EPA implemented water-quality standards. In a report dated April 29, 2008, the number of national impaired waters was 39,918 and the leading impairment was pathogens, totaling 9191 of reported impaired waters, or 14.16% (iaspub.epa.gov/waters/national_rept.control). A water body is defined as impaired when the water body fails to maintain water quality standards even after applying effluent limits for point sources (Clean Water Act Section 303(d); ww.epa.gov/waterscience/standards/303.htm). Water quality standards are based on water quality conditions and pollution sources, i.e., pathogens, nutrients, sediments, metals, habitat alteration and specific chemicals (www.epa.gov/waterscience/standards/about/).

The EPA Beaches Environmental Assessment and Coastal Health Program (BEACH) Act of 2000 was decreed to improve public health and recreational water quality. Fecal contaminated recreational waters may pose a potential health risk as bathers could contract a waterborne disease spread by fecal-oral route (www.epa.gov/waterscience/beaches/report/chapter02.pdf). Beach-goers who swim or bath in fecal-contaminated waters are at a greater risk than non-swimmers for contracting gastroenteritis. To protect public health, the use of fecal indicators as water quality standards were implemented (EPA, 1983; EPA, 1984). The BEACH Act (amended Section 303 of the Clean Water Act) requires states, tribes and territories to integrate EPA's water quality criteria, *E. coli* and/or enterococci, as their water quality standard (EPA, 2003). The focus of the BEACHs program is to strengthen beach testing and standards, provide faster testing

methods, predict pollution, to better define the criteria as to which fecal indicators and water quality standards are based, to invest in health and methods research and to inform the public (EPA, 2003). The BEACH Act requires states to adopt water quality standards that are "as protective of human health" as the federal criteria. The federal water quality criteria "Ambient Water Quality Criteria for Bacteria in Recreational Waters, 1986" was developed from US EPA epidemiology studies conducted over a period of years (1972 - 1978) at beaches located in New York, Louisiana, Massachusetts and Egypt (EPA, 1983; EPA, 1984). Epidemiological data supported the use of *E. coli* and enterococci as primary fecal indicators associated with statistically significant increased gastrointestinal illness rates to swimmers/bathers (EPA, 1983; EPA, 1984). A review of epidemiological studies and fecal indicators (Pruss, 1998) concluded the following: 1) an exposure-response relationship exists in recreational waters between bacterial indicator counts and gastrointestinal symptoms in exposed beach-goers (swimmers) and 2) there was no demonstrated relationship between bacterial indicator counts and non-gastrointestinal symptoms, i.e., rash, eyes, nose, ears.

The primary aim of the World Health Organization's (WHO) "Guidelines for Safe Recreational Water Environments" (2003) is to protect the public health by addressing such issues as exposure to sewage-contaminated waters, exposure to freeliving pathogenic organisms such as *Vibrio*, *Aeromonas* sp. and *N. fowleri*, exposure to contaminated beach sand and other potential exposures encountered in recreational water use. Unlike the EPA epidemiology studies conducted in the United States, the WHO based their selection of fecal water quality indicators and health-risk outcomes on a series of randomized control trials conducted in the United Kingdom (REF). The selected bacterial fecal indicator for marine

waters was enterococci whereby a dose-response relationship between enterococci density and health outcome, i.e., gastrointestinal illness and acute febrile respiratory illness (AFRI), was demonstrated. However, the WHO document did not recommend or find a statistical salient

fecal indicator for freshwaters. Enterococci are also the EPA fecal indicator for monitoring marine recreational waters and *E. coli* is an indicator in freshwaters.

Obstacles to current fecal indicator methods include bacterial culture-based methods, or if a molecular assay is used, the protocols usually involve bacterial or viral RNA/DNA concentration and purification steps thereby increasing the time frame between sample collection, data analysis, public health intervention and protection. Most culture-based detection methods currently require at least a 24-48 hr time lag from sample collection to outcome and therefore, provide information that fecal contamination occurred within the past 24-48 hr. Public health intervention to protect bathers prior to exposure would necessitate sample analysis and data confirmation to occur within hours of sample collection, not days. This lag time between sample collection, completion of analysis and public notification causes a window of potential risk to bathers from exposure to pathogens. Clearly, real-time or short-term detection, with limited (1-4 hr) turnaround time from water sample collection to results to public notification are imperative for timely protection of bathers.

Although both the EPA and WHO have developed rigorous recreational water quality guidelines, one limitation is that the bacterial indicators have little or no correlation to the presence of pathogenic viruses (Griffin et al., 2003). Bacterial indicators may be an erroneous predictor of viral presence as their survival rates do not match those of viruses. Even when

current bacterial standards are met in recreational waters, risks to human health may be posed by viruses. For example, most illnesses contracted by swimmers appear to be of viral etiology (Griffin et al., 2003). Erroneous bacterial counts have been documented as these fecal indicator bacteria periodically occurred naturally in temperate climates (Hardina and Fujioka, 1991; Roll and Fujioka, 1997; Byappanahalli and Fujioka, 1998; Fujioka and Byappanahalli, 2000; Solo-Gabriele HM et al., 2000; Genthner et al., 2005). Elevated bacterial indicator counts exceeding EPA water-quality criteria were influenced by soil run-off and not a result of sewage input (Byappanahalli and Fujioka, 2004).

To date, a viral indicator has not been mandated for regulatory purposes in recreational waters. Additional gaps for determining fecal contamination are that these methods are not real-time nor do they provide information regarding source. To minimize risks to human health, resource managers and human health advisors need an early-warning indicator, an indicator that addresses fecal source and an indicator indicative of enteric viral presence.

This attribute as to the selection of a fecal indicator(s) is based on the relationship between the indicator densities in polluted waters to human-health risk. Few epidemiology studies exists correlating health risks with male-specific (F+) coliphage densities. A California beach study comparing the male-specific ssRNA phage (FRNA) and male-specific DNA phage (FDNA) densities and gastrointestinal illness rates, nausea, cough and fever suggested an association between F+ phages and illness (Colford et al., 2007). Meta-analysis of epidemiological freshwater studies reported an elevated gastrointestinal illness risk with elevated bacteriophage exposure (Wade et al., 2003). However, the meta-analysis did not

specify which type of bacteriophage was evaluated.

In conjunction with microbial source-tracking, a tiered approach to rapid and effective detection and management of pathogen risks from fecal contamination encompasses a "big picture" investigation of a contaminated area. For example, Boehm et al., (2003) applied the tiered approach to resolve a closure of a local beach populated by tourists in the following manner: (1) determine the contamination source, i.e., runoff or wastewater discharge (2) survey the potential source area, i.e., beach gull droppings, broken sewer pipes, stormwater drains and (3) apply microbial library-independent microbial source tracking to water samples. Results of the above study suggested that a suite of indicator organisms and an established FRNA threshold value with a concomitant tier could potentially yield a more accurate, precise and timely environmental site assessment.

Methods developed in this study could also be used to address waterborne transmission of infectious diseases. For example the Severe Acute Respiratory Syndrome (SARS) was disseminated through aerosolized droplets of sewage (WHO, 2003). Had a viral indicator such as FRNA been detected during surveillance studies perhaps the sewage transmission via fecal droplets may have been detected sooner and thereby prevented or minimized the outbreak. The proposed viral indicator of recreational water quality, FRNA, and the molecular assays developed here could improve 1) monitoring criteria for Total Maximum Daily Loads (TMDL), state and federal recreational water quality regulations, ambient water monitoring programs, 2) combined microbiological-epidemiological studies designed to improve water quality criteria and 3) monitoring systems for wastewaters, biosolids, beach (recreational) waters, drinking waters, irrigation waters and reuse waters.

This dissertation project concluded the following: (1) analysis of complete genomic sequences from 30 *Leviviridae* FRNA coliphage strains plus two novel JS strains support the current classification scheme of two genera, *Levivirus* and *Allolevivirus* and four distinct genogroups, I, II, III and IV (2) FRNA sequences generated in this study will triple the genetic information currently available in the national genetic database for *Leviviridae* viruses (3) this is the first report of evidence for recombination in FRNA coliphages and (4) the genogroup-specific primer sets and RT-PCR amplification procedures should facilitate improved and reliable genotyping of FRNA coliphages.

IX. Summary and Conclusions

Environmental pressures, genetic mutation rates, microbial cross-species plasmid exchange, episodic outbreaks, global health threats, i.e., SARS and Avian Flu, and natural disasters such as floods and hurricanes along with an increasing societal coastal population contribute to the need for better indicator species. Public health microbiologists are challenged to address those needs.

Various wastewater treatment processes, environmental stressors and predation may influence selection of one indicator versus another indicator species. Therefore, a suite of microbial indicator species would add confidence to water-quality public health assessment.

FRNA phages are valuable models as surrogates for enteric viruses for the following reasons: (i) structure - similar icosahedral structure (ii) size - virion diameter (iii) morphology - are almost indistinguishable, by electron microscopy, to Picornoviruses, i.e.. poliovirus, enterovirus (iv) composition - most enterics contain ssRNA with the exception of Adenovirus (DNA) (v) site of replication - gastro-intestinal tract. FRNA and enteric viruses are both excreted by humans, FRNA are easily detectable and are approx 100X more abundant in wastewaters and raw sewage when compared to cytopathogenic enteric viruses (Grabow, 2001).

FRNA infect the host when bacteria are in log phase and under optimal conditions and temperatures whereas somatic phages infect a greater variety of bacteria genera, including

attachment to dead host cells. Somatic phages may replicate, under certain temperature and climate conditions, but it is very unlikely FRNA would replicate in the environment. Source-tracking of FDA phages is inconclusive although strain M13 occurs at approximately 77% in wastewaters. The presence of somatic, male-specific and *B. fragilis* phages in waters associated with animal and/or human wastes generally indicates the presence of enteric viruses and these phages potentially outnumber enteric viruses.

Rarely, if ever, does a direct correlation exist between the number of coliphages and enteric viruses at any given time. Enterics are excreted by infected humans whereas coliphages are excreted at all times. The incidence of human enterics in the form of outbreaks is seasonal and influenced by vaccination regimes. The excretion of coliphages is not affected by these occurrences.

Among various phages, FRNA are probably the best model for the presence of enteric viruses in the environment. Genogroup nucleotide and amino acid sequence alignment of FRNA phages originating from world-wide sources demonstrates the potential of FRNA to be applicable to various geographical locations and water sources.

FRNA sequences generated in this study tripled the genetic information currently available in the national genetic database for the *Leviviridae*. This additional genetic information may someday contribute to a better understanding of the basic molecular biology of these phages in terms of gene expression, control and regulation, recombination, mutations, virus-host interactions and phylogenetic relationships. In the more immediate future, this data can be applied to methods using FRNA coliphages as a fecal and viral indicator and as a source-tracking too

X. Recommendations for Future Research

Reverse-transcription PCR (RT-PCR) and real-time can be effectively used for source identification if specific primer sets are designed to be capable of identifying and distinguishing all members within each genogroup. Recently, reverse transcription polymerase chain reaction (RT-PCR) (Dryden et al., 2006) and reverse quantitative (RTQPCR) techniques (Kirs and Smith, 2007; Ogorzaly and Gantzer, 2006) for genotyping FRNA coliphages were developed. In these reports, primer design was based on the limited number of complete or partial nucleotide sequences available in GenBank. Thus, broad primer and probe genogroup specificity may not have been achieved due to the limited nucleotide data from the few strains available. To provide an advanced decision making tool for a comprehensive assessment of the sanitary quality of recreational water, a real-time RT-QPCR FRNA phage-based assay will be designed to allow for rapid detection and greater sensitivity. When developing real-time primers and TaqMan probes, guidelines for design included the absence of a G residue at the probe's 5' end, standard primer Tm should be 58-60 ^{o} C, probe should be 7-10 degrees higher than the primer set and the probe should be located close, just downstream from the forward primer. The large data set of 30 complete FRNA phage genomes will be used to align the members of all four genogroups. Therefore, optimal FRNA sequence target regions will be used to develop genogroup specific real-time primers and probes. Primer and probe sequences will be selected from the replicase gene for groups I

and IV and the capsid gene for groups II and III (Table 8.2). Tm's of cleavage probes will range from 66-70 ^{o} C, Tm's for the primers from 58-61^{o} C. Amplicons produced from all four primer sets will range from 99-153 bp in length. The method will incorporate a rapid heat-release procedure. To evaluate and control for false-negatives resulting from inhibition or other impurities leading to negative PCR outcomes, a non-competitive purified RNA (courtesy of Dr. Bill Burkhardt) will be incorporated as an internal RNA real-time control to account for false-negatives resulting from inhibition or other impurities leading to negative PCR outcomes. Variations in detection format, instrumentation, amplification efficiency, inhibition, various water matrix compositions, technical expertise and the step-wise process of real-time PCR will be accounted for by means of this internal control. This RNA molecule, approximately 300 nucleotides in length, possessed a unique pseudo-randomly generated ribonucleotide sequence. The primer set produces an amplicon of 149 bases. The probe will be labeled with Texas red with a Black Hole 1 quencher. Validation of this assay will begin in a phosphate buffer amended with raw sewage, representing a human source and/or feces from a variety of warm-blooded animals. Once validated with amended samples then water samples contaminated with known sources of fecal pollution will be used for the field validation.

In some recreational water samples FRNA coliphage densities would be below the limit of detection. When present in low numbers (< 5 X 10^2 PFU/ml) phages need to be enriched *in vivo.* Thus, an enrichment procedure will bring the densities of these FRNA phages above the limit of detection. Briefly, 18 ml of a recreation water sample will be placed in a 37 $^{\circ}$ C water bath. To the water sample, two ml of 10X TSB, streptomycin sulfate

(15 mg/L), ampicillin (15 mg/L) and 0.5 ml of an overnight culture of the host strain, *E. coli* HS (pFamp)R (DeBartolomeis and V.J. Cabelli, 1991) will be added. This suspension will be incubated for at least 90 min to allow the completion of one lytic cycle. Coliphage enrichment elevates the phage densities above the limit of detection as the burst size of each strain can range from between 2000 and 4500 PFU (Furuse, 1987). Therefore, the proportion of phages belonging to each genogroup may differ from what was present in the origin sample thereby introducing genogroup bias.

Although the one-step RT-PCR molecular assay designed in this study was based on full-length genomic sequences from 30 FRNA coliphages with a minimum of 5 to a maximum of 10 full-length genomic sequences/genogroup, more sequencing data is needed. In this study, phages were collected from ambient water sources, oysters, mussels, sewage and a few animal droppings (birds), swine lagoon or obtained from international collections. Future research would be to collect more phages from specific sources and, if possible, use a direct fecal swab or fresh voids. To prepare a large library of known FRNA coliphages and the respective source, collections need to be obtained directly from adults, children, pets and livestock and not mixtures of municipal sewage or livestock lagoons.

Limits of detection were not established in the traditional one-step RT-PCR, however, future plans are to perform limits of detection when developing the real-time PCR assay. To further develop the one-step RT-PCR, limits of detection will also be determined by first preparing log10 serial dilutions in phosphate buffered saline (PBS) using known amounts of plaque-forming units/ml (PFU/ml) as established with titer plates. At least two different strains/genogroup, for a total of 16 FRNA strains, will be titered and serial dilutions prepared.

From the respective serial dilution, each PCR reaction will be prepared using heat-released viral RNA followed by RT-PCR. When a PCR amplicon is no longer detected the limits of the assay will have been obtained.

Future studies will be to conduct a collaborative or round-robin performance evaluation of the RT-PCR and QRT-PCR methods developed in this study. In these performance evaluations multiple laboratories and technical staff subject the coliphage molecular assays to environmental samples and also evaluate "blinded" (positive control or reference) specimens in order to verify that it is possible for the labs and their analysts to correctly detect the FRNA coliphage genotype. Collaborative studies would serve to standardize the methods and would reveal any possible deficiencies or discrepancies that may have gone unnoticed during the initial assay development.

Short-term enrichments possibly followed by centrifugation, sterile filtration, heat-release and the use of PCR thermal cycles designed for short cycling regimes could very well lead to a rapid turnaround time from sample collection to results outcome for decision making. The use of heat-release of viral RNA templates alleviates the need for cumbersome and costly RNA purification by chemical methods. Ultimately, the molecular coliphage assay may be applicable to the development and eventual use of a field-portable kit whereby all reagents would be available in one tube. The field staff would only need to add an aliquot of an enriched water sample, heat and place into a portable thermal cycler.

Epidemiology field studies such as those of the current EPA Beaches program (National Epidemiological and Environmental Assessment of Recreational Water) could readily implement the FRNA RT-PCR or QPCR into their microbiological monitoring. Data

gathered in the EPA Beaches program is analyzed to evaluate exposed-bather health-risk associations with microbial water quality parameters. If FRNA coliphage quantification and genotyping is incorporated into the Beaches study, the data would provide epidemiological information on bather health impacts in relation to the concentrations of different coliphage genotypes in water, making it possible to associate FRNA presence and concentration thresholds with bather health effects.

The FRNA genome sequences provided by this dissertation research project are applicable to the development of an indicator-based microarray or hand-held fluorogenic detector assay. A comprehensive microarray constructed from viral and bacteria indicators and pathogens could potentially detect multiple pathogens and indicators and pin-point their sources with minimal laboratory lag time. Pathogen detection kits, similar to the AgPath-ID One-Step RT-PCR (Ambion) could be designed to detect FRNA genogroup targets. Results from such assays are generated in about 1 hr using a single-tube, real-time RT-PCR. The availability of the FRNA genetic database in conjunction with the primer and probes developed and evaluated in this dissertation provides the foundation for an assortment of rapid molecular detection assays.

Future applications of molecular FRNA coliphage assays include surveillance, monitoring and quality verification of groundwater and surface waters, shellfish beds, produce, red meat and poultry, assessments of wastewater treatment plant performance and effluent quality, wastewater reuse and surveillance for contamination incidents harboring emerging waterborne pathogens in sewage and ambient water bodies, surrogate for biological warfare and surrogate for enteric virus presence and control measures. In addition to current

bacterial indicators, FRNA coliphages have the potential to serve as effective viral indicators of fecal pollution.

References

Ashbolt NJ, WOK Grabow and M Snozzi. 2001. Indicators of microbial water quality. In: WHO Water Quality: Guidelines, Standards and Health.

Boehm, AB, JA Fuhrman, RD Mrse and SB Grant. 2003. Tiered approach for identification of a human fecal pollution source at a recreational beach - case study at Avalon Bay, Catalina Island, CA. Env Sci & Technol, 37(4):673-680.

Bosch A. 1998. Human enteric viruses in the water environment: a minireview. Internatl Microbiol 1:191-196.

Byappanahalli MN and RS Fujioka. 1998. Evidence that tropical sould environment can support the growth of Escherichia coli. Wat Sci Tech 38(12):171-174.

Cole D, SC Long and MD Sobsey. 2003. Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. Appl Environ Micro 69(11):6507-6514.

Colford JM, Jr., TJ Wade, KC Schiff, CC Wright, JF Griffith, SK Sandhu, S Burns, M Sobsey, G Lovelace and SB Weisberg. 2007. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology 18(1):27-35.

Dore WJ, K Henshilwood and DN Lees. 2000. Evaluation of F-specific RNA bacteriophage as a candidate human enteric virus indicator for bivalve molluscan shellfish. Appl Environ Micro 66(4):1280-1285.

Durán AE, M Muniesa, L Mocé-Llivina, G Campos, J Jofre and F Lucena. 2003. Usefulness of different groups of bacteriophages as model micro-organisms for evaluating chlorination. J Appl Microbiol, 95:29-37.

EPA. 1984. Health Effects Criteria for Fresh Recreational Waters. EPA-600/1-84-004.

EPA. 1983. Health Effects Criteria for Marine Recreational Waters. EPA-600/1-80-031.

EPA. 1986. Ambient Water Quality Criteria for Bacteria - 1986. EPA 440/5-84-002.

EPA. 2003. Bacterial water quality standards for recreational waters (freshwater and marine waters) status report. EPA-823-R-03-008. June, 2003.

Fujioka RS and MN Byappanahalli. 2000. Microbial ecology controls the establishment of

fecal bacteria in tropical soil environment. Proc Int Symp of the Center of Excellence, March 2000, pp 99-108. University of Tokyo and Minister of Education, Science, Sports and Culture of Japan.

Furuse K, S Osawa, J Kawashiro, R Tanaka, A Ozawa, S Sawamura, Y Yanagawa, T Nagao and I Watanabe. 1983. Bacteriophage distribution in human faeces: continuous survey of healthy subjects and patients with internal and leukaemic diseases. J Gen Virol, 64:2039-2043.

Furuse K. 1987. Distribution of coliphages in the environment: general considerations. In: *Phage Ecology*. SM Goyal, CP Gerba and G Britton, eds. John Wiley & Sons, Inc., NY. pp 87-124.

Genthner FJ, JB James, DF Yates and SD Friedman. 2005. Use of composite data sets for source-tracking enterococci in the water column and shoreline interstitial waters on Pensacola Beach, Florida. Mar Poll Bull 50:724-732.

Grabow, W. 2001. Bacteriophages: Update on application as models for viruses in water Water, SA 27: 251-268.

Gerba, CP. 2000. Assessment of enteric pathogen shedding by bathers during recreational activity and its impact on water quality. Quant Microbiol (2):55-68.

Gerba, CP. 1987. Phage as indicators of fecal pollution. In: *Phage Ecology*. SM. Goyal, CP. Gerba, and G. Britton (eds). John Wiley & Sons, Inc., NY. pp:197-209..

Griffin D, CJ Gibson III, EK Lipp, K Riley, JH Paul, and JB Rose. 1999. Detection of viral pathogens by reverse transcriptase PCR and of microbial indicators by standard methods in the canals of the Florida Keys. Appl Environ Micro 65:4118-4125.

Griffin DW, KA Donaldson, JH Paul and JB Rose. 2003. Pathogenic human viruses in coastal waters. Clin Micro Rev 16(1):129-143.

Hardina CM and RS Fujioka. 1991. Soil: the environmental source of *Escherichia coli* and enterococci in Hawaii's streams. Environ Tox Water Qual 6:185-195.

Leclerc H, S Edberg, V Pierzo, and JM Delattre. 2000. Bacteriophages as indicators of enteric viruses and public health risk in groundwater. J Appl Microbiol, 88(1):5-21.

Long SC, SS El-Khoury, SJG Oudejans, MD Sobsey, and J Vinjé. 2005. Assessment of sources and diversity of male-specific coliphages for source tracking. Env Eng Sci 22(3):367-377.

Nappier SP, MD Aiken and MD Sobsey. 2006. Male-specific coliphages as indicators of thermal inactivation of pathogens in biosolids. Appl Environ Microbiol, 72:2471-

2475.

Osawa S, K Furuse, and I Watanabe. 1981. Distribution of ribonucleic acid coliphages in animals. Appl Environ Microbiol 41(1):164-168.

Pruss A. 1998. Review of epidemiological studies on health effects from exposure to recreational water. Int J Epidemiol, 27:1-9.

Roll BM and RS Fujioka. 1997. Sources of fecal indicator bacteria in a brackish tropical stream and their impact of recreational water quality. Wat Sci Tech 35:179-186.

Schaper M, J Jofre, M Uys and WO Grabow. 2002. Distribution of genotypes of F-specific RNA bacteriophages in human and non-human sources of faecal pollution in South Africa and Spain. J Appl Micro 92(4):657-667.

Scott TM, JB Rose, TM Jenkins, SR Farrah and J Lukasik. 2002. Microbial source tracking: current methodology and future directions. Appl Environ Micro, 68:5796-5803.

Snow J. 1855. On the mode of communication of cholera. London: John Churchill, New Burlington Street, England.

Solo-Gabriele HM, MA Wolfert, TR Desmarais and CJ Palmer. 2000. Sources of *Escherichia coli* in coastal subtropical environment. Appl Environ Micro 66(1):230-237.

Stewart JR. 2002. Microbial source tracking using F+ RNA coliphage typing and Escherichia coli antibiotic resistance assays. Ph.D Dissertation. University of North Carolina at Chapel Hill, NC. School of Public Health, Department of Environmental Sciences and Engineering.

Vinjé J, SJG Oudejans, JR Stewart, MD Sobsey and SC Long. 2004. Molecular detection and genotyping of male-specific coliphages by reverse transcription-PCR and reverse line blot hybridization. Appl Environ Micro 70 (10):5996-6004.

World Health Organization. 2003. Emerging issues in water and infectious disease. 22 pp.

World Health Organization. 2003. Guidelines for safe recreational water environments. Vol 1. Coastal and Fresh Waters.

Alignment of Nucleotide Genomic Sequences
Appendix A1
Group I

Alignment: Levivirus Group I.

```
             ....|....| ....|....| ....|....| ....|....| ....|....|
                      10         20         30         40         50
Consensus    GGGTGGGACC CCTTTCGGGG TCCTGCTCRA CTTCCTGTCR AGCTAAATGC
DL1          .......... .......... .........A .........G ..........
DL2          ---------- ---------- ---------- ---------- ----------
DL13         ---------- ---------- ---------- ---------- ----------
DL16         .......... .......... .........A .........G ......-...
ST4          .......... .......... .........A .........G ......-...
R17          .......... .......... .........A .........G ......-...
J20          .......... .......... .........A .........G ......-...
MS2          .......... .......... .........A .........G ......-...
M12          .......... .......... .........A .........G ......-...
fr           .......... .......... .........G .........A ..........

             ....|....| ....|....| ....|....| ....|....| ....|....|
                      60         70         80         90        100
Consensus    CATTTTTAAT GTCTTTAGCG AGACGCTACC WTGGCTATCG CTGTAGGTAG
DL1          .......... .......... .......... A......... ..........
DL2          ---------- ---------- ----------. A......... ..........
DL13         ---------- ---------- ----------. A......... ..........
DL16         .......... .......... .......... A......... ..........
ST4          .......... .......... .......... A......... ..........
R17          .......... .......... .......... A......... ..........
J20          .......... .......... .......... A......... ..........
MS2          .......... .......... .......... A......... ..........
M12          .......... .......... .......... T......... ..........
fr           .......... .......... .......... A......... ..........

                                                      ORF1
             ....|....| ....|....| ....|....| ....|....| ....|....|
                     110        120        130        140        150
Consensus    CCGSAATTCC ATTSCTAGGR RGYYTSRYBY RYGMRAGYTY WYASYRHYCK
DL1          ...G...... ...C.....A G.TT.GACTC AT.CG..C.T TC.GTGTC.T
DL2          ...G...... ...C.....A G.TT.GACTC AT.CG..C.T TT.GTGTC.T
DL13         ...G...... ...C.....A G.TT.GACTC AT.CG..C.T TT.GTGTC.T
DL16         ...G...... ...C.....A G.TT.GACTC AT.CG..C.T TT.GTGTC.T
ST4          ...G...... ...C.....A G.TT.GACCT GT.CG..C.T TT.GTACC.T
R17          ...G...... ...C.....A G.TT.GACCT AT.CG..C.T TT.GTGCC.T
J20          ...G...... ...C.....A G.TT.GACTC AT.CG..C.T TT.GTGTC.T
MS2          ...G...... ...C.....A G.TT.GACCT GT.CG..C.T TT.GTACC.T
M12          ...G...... ...C.....A G.TT.GACTC AT.CG..C.T TT.GTGTT.T
fr           ...C...... ...G.....G A.CC.CGTGT GC.AA..T.C AT.CCAAC.G
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       WGAKMRKGAR WVYRARASCY WYGTSSYBHB CRTYCGCRHH TAYGCKRACG
DL1             T..TAAG..G TCTG.G.C.T TT..CCCGCT .G.C...ACC ..T..TG...
DL2             T..TCAG..G TCCG.G.C.T TC..CCCTAG .G.C...GTT ..T..GG...
DL13            T..TCAG..G TCCG.G.C.T TC..CCCTAG .G.C...GTT ..T..GG...
DL16            T..TCAG..G TCCG.G.C.T TC..CCCTAG .G.C...GTT ..T..GG...
ST4             T..TAGG..G AACG.G.C.T TC..CCCCTC .G.T...GTT ..C..GG...
R17             T..TAAG..G AGCA.G.C.T TC..CCCTTC .A.T...GTT ..C..GA...
J20             T..TAGG..A TCCG.G.C.T TC..CCCCTC .G.T...GTT ..T..TG...
MS2             T..TAGG..G AACG.G.C.T TC..CCCCTC .G.T...GTT ..C..GG...
M12             T..TCAG..G AACG.A.C.T TC..GCCCTC .G.T...GTT ..T..GG...
fr              A..GAAT..G --TA.G.G.C AC..GGTCTC .G.C...GAA ..T..TG...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       GBSARVYYGA RGATAACTCD TTHYCBYTVA WWTAYCGHTC SAAYTGGWCB
DL1             .CG.GGTT.. A........G ..CT.CT.G. AA..C..T.. G..C...A.C
DL2             .GC.GGTC.. A........G ..TT.CC.A. AA..C..T.. G..C...A.T
DL13            .GC.GGTC.. A........G ..TT.CC.A. AA..C..C.. G..C...A.T
DL16            .GC.GGTC.. A........G ..TT.CC.A. AA..C..C.. G..C...A.T
ST4             .TG.GACT.. A........A ..CT.TT.A. AA..T..T.. G..C...A.T
R17             .TG.GACC.. A........A ..CT.TT.A. AA..T..C.. G..C...A.T
J20             .GG.GGTC.. G........A ..TT.CT.A. AA..C..T.. G..C...A.T
MS2             .TG.GACT.. A........A ..CT.TT.A. AA..T..T.. G..C...A.T
M12             .CG.GACT.. A........G ..TT.CC.C. AA..T..C.. G..C...A.C
fr              .GG.ACTC.. G........T ..AC.GT.G. TT..C..A.. C..T...T.G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       CCBGGYCRDT WYAMHWSKAC BGGKNCBMRM ACRRADSART GGCACTAYCC
DL1             ..T..T.GT. TT.ACTCG.. T..GG.TAGA ..GA.AC.G. .......T..
DL2             ..C..T.GA. TT.ATTCG.. C..GT.TAGA ..GA.AC.G. .......T..
DL13            ..C..T.GA. TT.ATTCG.. C..GT.TAGA ..GG.AC.G. .......T..
DL16            ..C..T.GA. TT.ATTCG.. C..GT.TAGA ..GG.AC.G. .......T..
ST4             ..C..T.GT. TT.ACTCG.. T..GG.CAAA ..GA.AC.G. .......C..
R17             ..C..T.GT. TT.ACTCG.. T..GG.CAGA ..GA.GC.G. .......T..
J20             ..T..C.GA. TT.ACTCG.. T..GA.CAGA ..GA.TC.G. .......T..
MS2             ..C..T.GT. TT.ACTCG.. T..GG.CAAA ..GA.AC.G. .......C..
M12             ..G..T.GA. TT.ACTCG.. T..GG.CAGA ..GA.AC.G. .......T..
fr              ..G..C.AG. AC.CAAGT.. G..TC.GCGC ..AA.GG.A. .......C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       VTCBYCBTAY TCDMGDGGDG CGHTNRGHRT YAMDKCKVTN GATCAAGGTD
DL1             G..CT.T..T ..TA.A..T. ..C.TA.CG. C.CAT.GG.G .........G
DL2             G..CC.T..T ..TA.G..T. ..C.CA.TG. C.CAT.GG.A .........G
DL13            G..CC.T..T ..TA.G..T. ..C.CA.TG. C.CAT.GG.A .........G
DL16            G..CC.T..T ..TA.G..T. ..C.CA.TG. C.CAT.GG.A .........G
ST4             C..TC.G..T ..AC.G..G. ..T.AA.TG. C.CAT.GA.A .........G
R17             C..CC.G..T ..GC.G..G. ..T.AA.TG. C.CGT.GA.A .........G
J20             G..CC.T..C ..TA.G..A. ..C.TA.TG. C.CTT.GG.C .........T
MS2             C..TC.G..T ..AC.G..G. ..T.AA.TG. C.CAT.GA.A .........G
```

```
M12          C..CC.T..C ..TC.G..A. ..T.GA.TG. T.CTG.GA.A .........G
fr           A..GT.C..C ..AC.T..G. ..A.AG.AA. C.AGG.TC.T .........A
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       MBTAYRMRCG MWBKGGSWCR TCGTGGGGYC GYSMGTWCGA RGARMRWRCY
DL1             CT..TAAG.. CTCT..GT.A ........T. .CCC..A... G..GAAAG.C
DL2             CC..CAAG.. CTCT..GT.A ........T. .CCC..A... G..GAAAG.C
DL13            CC..CAAG.. CTCT..GT.A ........T. .CCC..A... G..GAAAG.C
DL16            CC..CAAG.. CTCT..GT.A ........T. .CCC..A... G..GAAAG.C
ST4             CC..CAAG.. AAGT..GT.A ........T. .CCC..A... G..GAAAG.C
R17             CC..TAAG.. CAGT..GT.A ........T. .CCC..A... G..GAAAG.C
J20             CC..CAAG.. CTCT..GT.A ........T. .TCC..A... G..GAAAG.T
MS2             CC..CAAG.. AAGT..GT.A ........T. .CCC..A... G..GAAAG.C
M12             CC..TAAG.. AAGT..GT.G ........T. .CCC..A... G..GAAAA.C
fr              AG..CGCA.. ATTG..CA.A ........C. .CGA..T... A..ACGTG.C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       GGTTWYGGYW TSTCNMTCGA CGCACGYWSY TGYTAYAGCC TMTTCCCYGT
DL1             ....AT..CT .C..AC.... ......CTCC ..C..C.... .C.....T..
DL2             ....TT..CT .C..AC.... ......TTCC ..C..T.... .C.....T..
DL13            ....TT..CT .C..AC.... ......TTCC ..C..T.... .C.....T..
DL16            ....TT..CT .C..AC.... ......TTCC ..C..T.... .C.....T..
ST4             ....TC..CT .C..CC.... ......CTCC ..C..C.... .C.....T..
R17             ....TT..CT .C..TC.... ......CTCC ..C..C.... .C.....T..
J20             ....TT..CT .C..GC.... ......TTCC ..C..C.... .C.....T..
MS2             ....TC..CT .C..CC.... ......CTCC ..C..C.... .C.....T..
M12             ....TT..TT .C..AC.... ......CTCC ..C..C.... .C.....T..
fr              ....AT..CA .G..TA.... ......TAGT ..T..T.... .A.....C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       HAGYCARAAY HTRACWTRSA THGAMGTRCC RMMGAACGTW GCKAAYCGSG
DL1             T..T..G..C C.G..T.AC. .T..A..G.. GCA......T ..G..T..G.
DL2             T..T..G..T C.G..T.AC. .C..A..G.. GCA......T ..G..C..G.
DL13            T..T..G..T C.G..T.AC. .C..A..G.. GCA......T ..G..C..G.
DL16            T..T..G..T C.G..T.AC. .C..A..G.. GCA......T ..G..C..G.
ST4             A..C..G..C T.G..T.AC. .C..A..G.. GCA......T ..G..C..G.
R17             A..C..G..C T.G..T.AC. .C..A..G.. GCA......T ..G..T..G.
J20             C..T..A..T A.G..T.AC. .C..A..G.. GCA......T ..G..T..G.
MS2             A..C..A..C T.G..T.AC. .C..A..G.. GCA......T ..G..C..G.
M12             A..C..G..C C.G..T.AC. .C..A..G.. GCA......T ..G..C..G.
fr              T..T..A..C T.A..A.GG. .A..C..A.. AAC......A ..T..T..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus       CBWCGACYGA RGTCYTRSRD AAGGTYACYC ARGGNAAYTT YAACCTTGGB
DL1             .GT....C.. A...C.GCAG .....C..C. .A..C..T.. C........T
DL2             .GT....C.. A...C.GCAG .....C..T. .A..A..T.. C........T
DL13            .GT....C.. A...C.GCAG .....C..T. .A..A..T.. C........T
DL16            .GT....C.. A...C.GCAG .....C..T. .A..A..T.. C........T
ST4             .GT....C.. A...C.GCAA .....C..C. .G..T..T.. T........T
R17             .GT....C.. A...C.GCAA .....T..C. .A..T..T.. C........T
J20             .GT....C.. A...C.GCAG .....T..C. .G..T..T.. T........C
MS2             .GT....C.. A...C.GCAA .....C..C. .G..T..T.. T........T
```

```
M12          .TA....C.. A...C.GCAG .....T..C. .A..G..C.. C........G
fr           .CA....T.. G...T.AGGT .....C..T. .A..T..T.. T........C
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       560        570        580        590        600
Consensus       GTNGCYYTWG CWGARGCVMG RTCKACRGCC TCACAACTSK CGACGCAAAC
DL1             ..A..CT.A. .A..G..GA. G..G..A... ........CG ..........
DL2             ..G..CC.A. .A..G..CA. G..G..A... ........CG ..........
DL13            ..G..CC.A. .A..G..CA. G..G..A... ........CG ..........
DL16            ..G..CC.A. .A..G..CA. G..G..A... ........CG ..........
ST4             ..T..TT.A. .A..G..CA. G..G..A... ........CG ..........
R17             ..C..CC.A. .A..G..CA. A..G..A... ........CG ..........
J20             ..A..TT.A. .A..G..CA. G..G..A... ........CG ..........
MS2             ..T..TT.A. .A..G..CA. G..G..A... ........CG ..........
M12             ..G..CC.T. .A..G..AA. G..G..A... ........CG ..........
fr              ..G..CC.T. .T..A..CC. G..T..G... ........GT ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       610        620        630        640        650
Consensus       CATYGCGYTS RTKAAGGCGT ACACYGCSGC TCGYCGCGGB AAYTGGCGCC
DL1             ...T...C.C G.G....... ....T..C.. ...T.....C ..C.......
DL2             ...T...C.C G.G....... ....T..C.. ...T.....C ..T.......
DL13            ...T...C.C G.G....... ....T..C.. ...T.....C ..T.......
DL16            ...T...C.C G.G....... ....T..C.. ...T.....C ..T.......
ST4             ...T...C.C G.G....... ....T..C.. ...T.....T ..T.......
R17             ...T...C.C G.G....... ....T..C.. ...T.....T ..T.......
J20             ...T...C.C G.G....... ....T..C.. ...T.....T ..T.......
MS2             ...T...C.C G.G....... ....T..C.. ...T.....T ..T.......
M12             ...T...C.C G.G....... ....T..C.. ...T.....T ..T.......
fr              ...C...T.G A.T....... ....C..G.. ...C.....G ..C.......

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       660        670        680        690        700
Consensus       AGVCRSTCCG CTAYYTHGCS CTDAACGARR AYCGRAARTT YMRDTCRAAR
DL1             ..A.GC.... ...TC.C..C ..T.....AG .C..A..A.. TCGG..A..A
DL2             ..A.GC.... ...TC.T..C ..A.....AG .T..G..G.. CCGA..G..A
DL13            ..A.GC.... ...TC.T..C ..A.....AG .T..G..G.. CCGA..G..A
DL16            ..A.GC.... ...TC.T..C ..A.....AG .T..G..G.. CCGA..G..A
ST4             ..G.GC.... ...CC.T..C ..A.....AG .T..A..G.. TCGA..A..A
R17             ..G.GC.... ...CC.C..C ..A.....AG .T..A..A.. TCGA..A..A
J20             ..G.GC.... ...TC.T..C ..A.....AG .T..G..A.. CCGG..A..A
MS2             ..G.GC.... ...CC.T..C ..A.....AG .T..A..G.. TCGA..A..A
M12             ..C.GG.... ...CC.C..C ..A.....AG .T..A..G.. TCGA..A..A
fr              ..G.AC.... ...CT.A..G ..G.....GA .T..A..A.. CAAT..G..G

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       710        720        730        740        750
Consensus       YMCGTSGCVR GYAGRTGGYT GGAGTTGCAG TTCGGNTGGH TRCCRCTHMT
DL1             CA...G..GG .C..G...T. .......... .....C...T .A..G..CA.
DL2             CA...G..AG .C..G...T. .......... .....C...C .A..G..CA.
DL13            CA...G..AG .C..G...T. .......... .....C...C .A..G..CA.
DL16            CA...G..AG .C..G...T. .......... .....C...C .A..G..CA.
ST4             CA...G..CG .C..G...T. .......... .....T...T .A..A..AA.
R17             CA...G..CG .C..G...T. .......... .....A...T .A..A..AA.
J20             CA...G..AG .T..A...T. .......... .....T...C .A..G..TA.
MS2             CA...G..CG .C..G...T. .......... .....T...T .A..A..AA.
```

```
M12          CA...G..CG .C..G...T. .......... .....C...T .A..A..TA.
fr           TC...C..AA .C..G...C. .......... .....G...A .G..G..TC.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        760        770        780        790        800
Consensus       SAGYGATATC CARGGYGCRT AYGAGATGCT YACSAARGTK CAYCTTMARG
DL1             G..C..... ..G..C..A. .T........ T..G..G..T ..C...C.A.
DL2             G..C..... ..G..C..A. .T........ T..G..G..T ..C...C.A.
DL13            G..C..... ..G..C..A. .T........ T..G..G..T ..C...C.A.
DL16            G..C..... ..G..C..A. .T........ T..G..G..T ..C...C.A.
ST4             G..T..... ..G..T..A. .T........ T..G..G..T ..C...C.A.
R17             G..T..... ..A..T..A. .T........ T..G..G..T ..C...C.A.
J20             G..C..... ..A..T..A. .T........ T..G..G..T ..T...C.A.
MS2             G..T..... ..G..T..A. .T........ T..G..G..T ..C...C.A.
M12             G..T..... ..G..C..A. .C........ T..G..G..T ..C...C.A.
fr              C..C..... ..A..T..G. .T........ C..C..A..G ..T...A.G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        810        820        830        840        850
Consensus       MRTTTMTBCC TATGMGDGCC GTRMGNCARG TNGGHMMWAA CRTYARKTTR
DL1             AG...C.C.. ....A.A... ..GC.C..G. .A..CACT.. .G.C.AG..A
DL2             AG...C.T.. ....A.G... ..AC.G..A. .G..CACT.. .A.T.AG..A
DL13            AG...C.T.. ....A.G... ..AC.G..A. .G..CACT.. .A.T.AG..A
DL16            AG...C.T.. ....A.G... ..AC.G..A. .G..CACT.. .A.T.AG..A
ST4             AG...C.T.. ....A.A... ..AC.T..G. .C..TACT.. .A.C.AG..A
R17             AG...C.T.. ....A.A... ..AC.T..A. .T..TACT.. .A.T.AG..A
J20             AG...C.T.. ....A.A... ..AC.A..G. .G..TACT.. .G.T.AG..A
MS2             AG...C.T.. ....A.A... ..AC.T..G. .C..TACT.. .A.C.AG..A
M12             AG...C.C.. ....A.A... ..AC.T..G. .C..CACT.. .A.T.AG..A
fr              CA...A.G.. ....C.T... ..GA.G..A. .C..ACAA.. .G.C.GT..G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        860        870        880        890        900
Consensus       DMTGGCCGBY TBDCKTMTCC RGCTGCAARC TWYMARWCWA CGTGCAACAT
DL1             GA......CC .GT.G.A... A.......A. .ACC.AA.T. ..........
DL2             GA......TT .GT.G.A... A.......A. .ACC.GA.T. ..........
DL13            GA......TT .GT.G.A... A.......A. .ACC.GA.T. ..........
DL16            GA......TT .GT.G.A... A.......A. .ACC.GA.T. ..........
ST4             AA......TC .GT.G.A... A.......A. .TCC.GA.A. ..........
R17             GA......CT .GG.G.A... A.......A. .TCC.GA.A. ..........
J20             GA......CT .GT.G.A... A.......A. .ACC.GA.T. ..........
MS2             GA......TC .GT.G.A... A.......A. .TCC.GA.A. ..........
M12             GA......CC .TT.G.A... A.......A. .ACC.GA.A. ..........
fr              TC......GC .CA.T.C... G.......G. .ATA.GT.T. ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        910        920        930        940        950
Consensus       ATCRCGACGH ATYGTGATAT GGTTTTACAT AAACGATGCA CGWYTGGCHT
DL1             ...G.....T ..C..... .......... .......... ..TT....T.
DL2             ...A.....A ..C..... .......... .......... ..TT....C.
DL13            ...A.....A ..C..... .......... .......... ..TT....C.
DL16            ...A.....A ..C..... .......... .......... ..TT....C.
ST4             ...G.....T ..C..... .......... .......... ..TT....A.
R17             ...A.....T ..C..... .......... .......... ..TT....A.
J20             ...G.....A ..C..... .......... .......... ..TT....C.
MS2             ...G.....T ..C..... .......... .......... ..TT....A.
```

```
M12          ...G.....T ..C....... .......... .......... ..AT....C.
fr           ...A.....C ..T....... .......... .......... ..TC....T.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       960        970        980        990       1000
Consensus       GGYTGTCSTC YYTRGGKATY TTGAACCCRC TAGGWATAGT GTGGGAAAAG
DL1             ..T....G.. TC.A..T..C ........A. ....T..... ..........
DL2             ..T....G.. TC.A..T..C ........A. ....T..... ..........
DL13            ..T....G.. TC.A..T..C ........A. ....T..... ..........
DL16            ..T....G.. TC.A..T..C ........A. ....T..... ..........
ST4             ..T....G.. TC.A..T..C ........A. ....T..... ..........
R17             ..T....G.. TC.A..T..C ........A. ....T..... ..........
J20             ..T....G.. TC.A..T..C ........A. ....T..... ..........
MS2             ..T....G.. TC.A..T..C ........A. ....T..... ..........
M12             ..T....G.. TC.G..T..C ........A. ....T..... ..........
fr              ..C....C.. CT.A..G..T ........G. ....A..... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      1010       1020       1030       1040       1050
Consensus       GTSCCBTTCT CWTTCSTKGT CGAYTGGYTS CTDCCKGTDG GDAACATGCT
DL1             ..G..T.... .A...G.T.. ...C...C.C ..T..T..A. .A........
DL2             ..G..T.... .A...G.T.. ...C...C.C ..T..T..T. .G........
DL13            ..G..T.... .A...G.T.. ...C...C.C ..T..T..T. .G........
DL16            ..G..T.... .A...G.T.. ...C...C.C ..T..T..T. .G........
ST4             ..G..T.... .A...G.T.. ...C...C.C ..A..T..A. .T........
R17             ..G..T.... .A...G.T.. ...C...C.C ..A..T..A. .T........
J20             ..G..T.... .A...G.T.. ...C...C.C ..T..T..A. .A........
MS2             ..G..T.... .A...G.T.. ...C...C.C ..A..T..A. .T........
M12             ..G..C.... .A...G.T.. ...C...C.C ..G..T..G. .G........
fr              ..C..G.... .T...C.G.. ...T...T.G ..G..G..T. .G........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      1060       1070       1080       1090       1100
Consensus       HGAGGGSCTH ACVGCYCCVR TDGGMTGYTC BTAYMWRTCD GGRACMGTWA
DL1             C.....C..A ..G..C..CG .A..A..T.. T..CATG..G ..A..A..T.
DL2             C.....C..C ..A..C..CG .A..A..C.. C..CATG..T ..A..A..T.
DL13            C.....C..C ..A..C..CG .A..A..C.. C..CATG..T ..A..A..T.
DL16            C.....C..C ..A..C..CG .A..A..C.. C..CATG..T ..A..A..T.
ST4             C.....C..T ..G..C..CG .G..A..C.. C..CATG..A ..A..A..T.
R17             C.....C..T ..G..T..CG .T..A..C.. C..CATG..A ..A..A..T.
J20             C.....C..T ..A..C..CG .A..A..C.. C..CATG..T ..A..A..T.
MS2             C.....C..T ..G..C..CG .G..A..C.. C..CATG..A ..A..A..T.
M12             A.....C..T ..G..T..AG .A..A..T.. T..CATG..A ..G..A..T.
fr              T.....G..T ..C..C..GA .A..C..T.. G..TCAA..G ..A..C..A.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      1110       1120       1130       1140       1150
Consensus       CYGACGTAAT AWCRGGWGAG TCSAYMATAA SCGYYGAYGM YMYCTAYGGK
DL1             .T........ .A.G..T... ..C.TC.... G..TT..C.C TCC...T..G
DL2             .T........ .A.G..T... ..C.TC.... G..TC..C.C TCC...T..G
DL13            .T........ .A.G..T... ..C.TC.... G..TC..C.C TCC...T..G
DL16            .T........ .A.G..T... ..C.TC.... G..TC..C.C TCC...T..G
ST4             .T........ .A.G..T... ..C.TC.... G..TT..C.C TCC...C..G
R17             .T........ .A.G..T... ..C.TC.... G..TT..C.C TCC...T..G
J20             .T........ .A.G..T... ..C.TC.... G..TT..C.C TCC...T..G
MS2             .T........ .A.G..T... ..C.TC.... G..TT..C.C TCC...C..G
```

```
M12          .T........ .A.G..T... ..C.TC.... G..TT..C.C TCC...C..G
fr           .C........ .T.A..A... ..G.CA.... C..CC..T.A CAT...T..T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1160       1170       1180       1190       1200
Consensus       TGGRMDRYRG WKMGACMKGS MACYGCTAAG GYSCADRTYW SDGCYRTSCA
DL1             ...ACAGTG. AGA...AG.G C..T..... .CC..GA.CT CA..CA.G..
DL2             ...ACGGTG. ATA...AG.G C..T..... .CC..AG.CT CA..CA.G..
DL13            ...ACGGTG. ATA...AG.G C..T..... .CC..AG.CT CA..CA.G..
DL16            ...ACGGTG. ATA...AG.G C..T..... .CC..AG.CT CA..CA.G..
ST4             ...ACTGTG. AGA...AG.G C..T..... .CC..AA.CT CA..CA.G..
R17             ...ACTGTG. AGA...AG.G C..T..... .CC..TG.TT CA..CA.G..
J20             ...ACTGTG. AGA...AG.G C..T..... .CC..GA.CT CG..CA.G..
MS2             ...ACTGTG. AGA...AG.G C..T..... .CC..AA.CT CA..CA.G..
M12             ...ACTGTG. AGA...AG.G C..T..... .CC..AG.CT CA..CA.G..
fr              ...GATACA. TGC...CT.C A..C..... .TG..AA.CA GT..TG.C..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1210       1220       1230       1240       1250
Consensus       YCGRGGGGTR CARWSCGTRT GSCCMACWAC KGGCGYRTAC GTDAARTCWC
DL1             T..A.....A ..GTC...A. .G..A..A.. T....TA... ..A..G..A.
DL2             T..A.....G ..ATC...A. .G..A..A.. T....TA... ..G..G..A.
DL13            T..A.....G ..ATC...A. .G..A..A.. T....TA... ..G..G..A.
DL16            T..A.....G ..ATC...A. .G..A..A.. T....TA... ..G..G..A.
ST4             T..A.....A ..ATC...A. .G..A..A.. T....CG... ..A..G..T.
R17             T..A.....A ..ATC...A. .G..A..A.. T....CA... ..A..G..T.
J20             T..A.....G ..ATC...A. .G..A..A.. T....TA... ..G..A..A.
MS2             T..A.....A ..ATC...A. .G..A..A.. T....CG... ..A..G..T.
M12             T..A.....A ..ATC...A. .C..A..T.. T....TA... ..A..G..T.
fr              C..G.....A ..AAG...G. .G..C..A.. G....TA... ..T..G..A.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1260       1270       1280       1290       1300
Consensus       CYTTYTCGAT KGTCCAYACY TTAGAYGCST TRGCAYTWWT CAGGCAACGS
DL1             .T..C..... G.....C..T .....T..G. .G...T.AA. .........G
DL2             .T..C..... G.....T..C .....C..G. .G...T.AA. .........G
DL13            .T..C..... G.....T..C .....C..G. .G...T.AA. .........G
DL16            .T..C..... G.....T..C .....C..G. .G...T.AA. .........G
ST4             .T..C..... G.....T..C .....T..G. .A...T.AA. .........G
R17             .T..C..... G.....T..C .....T..G. .A...T.AA. .........G
J20             .T..C..... T.....T..C .....T..G. .A...T.AA. .........G
MS2             .T..C..... G.....T..C .....T..G. .A...T.AA. .........G
M12             .C..T..... G.....T..C .....T..G. .G...T.AA. .........G
fr              .T..C..... G.....T..C .....T..C. .G...C.TT. .........C
                         STOP 1                                ORF2
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1310       1320       1330       1340       1350
Consensus       CTCTBDARAT AGAGSCCYYA ACCGRAGKKH GARCCRCATG GCTTCKAACT
DL1             ....CA.G.. ....C..TC. ....G..TTC ..A--G.... .....T....
DL2             ....TA.G.. ....C..TC. ....G..TTC ..A--G.... .....T....
DL13            ....TA.G.. ....C..TC. ....G..TTC ..A--G.... .....T....
DL16            ....TA.G.. ....C..TC. ....G..TTC ..A--G.... .....T....
ST4             ....CT.G.. ....C..TC. ....G..TTT ..A--G.... .....T....
R17             ....CT.A.. ....C..TC. ....G..TTT ..A--G.... .....T....
J20             ....CA.G.. ....C..TC. ....G..TTT ..A--G.... .....T....
MS2             ....CT.G.. ....C..TC. ....G..TTT ..A--G.... .....T....
M12             ....CT.A.. ....C..TC. ....G..TTT ..A--G.... .....T....
```

```
fr          ....GG.A.. ....G..CT. ....A..GGA ..G..A.... .....G....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1360       1370       1380       1390       1400
Consensus       TTRMWSAGTT YGTTCTCGTC GACAATGGCG GAACBGGHGA YGTRAMWGTC
DL1             ..ACTC.... C......... .......... ....T..T.. C..G.CT...
DL2             ..ACTC.... C......... .......... ....T..C.. C..G.CT...
DL13            ..ACTC.... C......... .......... ....T..C.. C..G.CT...
DL16            ..ACTC.... C......... .......... ....T..C.. C..G.CT...
ST4             ..ACTC.... C......... .......... ....T..C.. C..G.CT...
R17             ..ACTC.... T......... .......... ....T..C.. C..G.CT...
J20             ..ACTC.... T......... .......... ....C..C.. C..G.CT...
MS2             ..ACTC.... C......... .......... ....T..C.. C..G.CT...
M12             ..ACTC.... T......... .......... ....T..C.. C..G.CT...
fr              ..GAAG.... C......... .......... ....G..A.. T..A.AA...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1410       1420       1430       1440       1450
Consensus       GCBCCRAGCA ACTTCGCTAA CGGGGTYGCW GAATGGATYA GCTCBAACTC
DL1             ..C..A.... .......... ......C..T ........T. ....T.....
DL2             ..T..A.... .......... ......C..T ........T. ....T.....
DL13            ..T..A.... .......... ......C..T ........T. ....T.....
DL16            ..T..A.... .......... ......C..T ........T. ....T.....
ST4             ..C..A.... .......... ......C..T ........C. ....T.....
R17             ..T..A.... .......... ......C..T ........C. ....T.....
J20             ..T..A.... .......... ......C..T ........C. ....T.....
MS2             ..C..A.... .......... ......C..T ........C. ....T.....
M12             N.G..A.... .......... ......C..T ........C. ....C.....
fr              ..T..G.... .......... ......T..A ........C. ....G.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1460       1470       1480       1490       1500
Consensus       VCGYTCWCAR GCTTACAAAG TRACYTGTAG YGTKCGTCAG AGCTCTGCGM
DL1             G..C..T..G .......... .A..C..... C..T...... .........C
DL2             A..T..T..G .......... .A..C..... C..T...... .........C
DL13            A..T..T..G .......... .A..C..... C..T...... .........C
DL16            A..T..T..G .......... .A..C..... C..T...... .........C
ST4             G..T..A..G .......... .A..C..... C..T...... .........C
R17             G..C..A..G .......... .A..C..... C..T...... .........C
J20             A..C..T..G .......... .A..C..... T..T...... .........C
MS2             G..T..A..G .......... .A..C..... C..T...... .........C
M12             C..C..A..A .......... .A..T..... C..G...... .........C
fr              A..T..T..G .......... .G..C..... C..G...... .........A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1510       1520       1530       1540       1550
Consensus       ASAAYCGSAA RTACACYRTY AARGTYGARG TRCCDAARGT GGCWACYCAR
DL1             .G..T..C.. G.....TA.C ..G..C..A. .G..G..A.. ...T..C..G
DL2             .G..C..C.. G.....CA.T ..G..T..G. .A..A..A.. ...T..C..A
DL13            .G..C..C.. G.....CA.T ..G..T..G. .A..A..A.. ...T..C..A
DL16            .G..C..C.. G.....CA.T ..G..T..G. .A..A..A.. ...T..C..A
ST4             .G..T..C.. A.....CA.C ..A..C..G. .G..T..A.. ...A..C..G
R17             .G..T..C.. A.....CA.T ..A..C..G. .G..T..G.. ...A..T..G
J20             .G..C..C.. G.....CA.C ..A..C..A. .G..A..A.. ...T..T..G
MS2             .G..T..C.. A.....CA.C ..A..C..G. .G..T..A.. ...A..C..G
```

```
M12          .G..C..C.. A.....CA.C ..G..C..G. .G..G..A.. ...A..C..A
fr           .C..T..G.. A.....CG.C ..G..C..G. .G..G..A.. ...A..T..G
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       RYYSWHGGYG GYGTWSAGCT TCCTGTWGCS GCRTGGCGYT CGTAYHTRAA
DL1             ACCGTT..T. .C..AC.... ......A..C ..G.....T. ....TC.G..
DL2             ACCGTC..T. .T..AC.... ......A..C ..A.....T. ....CT.G..
DL13            ACCGTC..T. .T..AC.... ......A..C ..A.....T. ....CT.G..
DL16            ACCGTC..T. .T..AC.... ......A..C ..A.....T. ....CT.G..
ST4             ACTGTT..T. .T..AG.... ......A..C ..A.....T. ....CT.A..
R17             ACTGTT..T. .T..AG.... ......A..C ..A.....T. ....CT.A..
J20             ACCGTC..T. .C..AC.... ......A..C ..A.....T. ....CT.A..
MS2             ACTGTT..T. .T..AG.... ......A..C ..A.....T. ....CT.A..
M12             ACCGTT..C. .T..AG.... ......A..C ..A.....T. ....CC.G..
fr              GTCCAA..C. .C..TG.... ......T..G ..G.....C. ....CA.G..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1610       1620       1630       1640       1650
Consensus       TATGGAAYTR ACYATTCCDR THTTCGCDAC GAAYKMCGAY TGYGMSYTWA
DL1             .......T.G ..T.....AA .C.....A.. ...CGA...C ..C.CGC.A.
DL2             .......T.A ..T.....TA .T.....A.. ...CGA...T ..C.CGC.T.
DL13            .......T.A ..T.....TA .T.....A.. ...CGA...T ..C.CGC.T.
DL16            .......T.A ..T.....TA .C.....A.. ...CGA...T ..C.CGC.T.
ST4             .......C.A ..C.....AA .T.....T.. ...TTC...C ..C.AGC.T.
R17             .......T.A ..T.....AA .T.....T.. ...CTC...T ..C.AGC.T.
J20             .......T.A ..T.....AA .T.....G.. ...CGA...T ..C.CGC.T.
MS2             .......C.A ..C.....AA .T.....T.. ...TTC...C ..C.AGC.T.
M12             .......T.A ..T.....GA .T.....T.. ...CTC...T ..C.CGC.T.
fr              .......T.A ..T.....GG .A.....G.. ...CGA...C ..T.CCT.A.
                                                  ORF3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1660       1670       1680       1690       1700
Consensus       TYGTYAARGC RWTGCAAGGY MYCYTDAAAR MTGGWAACCC RATYSCHWCR
DL1             .T..T..G.. GA.......T CT.C.A...G A...A..... G..CC.CT.A
DL2             .T..T..G.. GA.......T CT.C.A...G A...A..... G..CC.CT.A
DL13            .T..T..G.. GA.......T CT.C.A...G A...A..... G..CC.CT.A
DL16            .T..T..G.. GA.......T CT.C.A...G A...A..... G..CC.CT.A
ST4             .T..T..G.. AA.......T CT.C.A...G A...A..... G..TC.CT.A
R17             .T..T..G.. AA.......T CT.C.A...G A...A..... G..TC.CT.A
J20             .T..T..G.. GA.......T CT.C.A...G A...A..... A..CC.CT.A
MS2             .T..T..G.. AA.......T CT.C.A...G A...A..... G..TC.CT.A
M12             .T..C..A.. AA.......T CT.C.G...G A...A..... G..TC.TT.G
fr              .C..T..G.. AT.......C AC.T.T...A C...T..... A..TG.AA.A
                                                  STOP 2
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1710       1720       1730       1740       1750
Consensus       GCMATCGCAG CMAACTCSGG MATCTAYTAA KARAYNYSKK CCATTCMAAC
DL1             ..A....... .A.....C.. A.....C... T.G.TTTCGG ......A...
DL2             ..A....... .A.....C.. A.....C... T.G.TTCCGT ......C...
DL13            ..A....... .A.....C.. A.....C... T.G.TTCCGT ......C...
DL16            ..A....... .A.....C.. A.....C... T.G.TTCCGT ......C...
ST4             ..A....... .A.....C.. C.....C... T.G.CGCCGG ......A...
R17             ..A....... .A.....C.. C.....C... T.G.TGCCGG ......A...
J20             ..A....... .C.....C.. A.....T... T.G.TTCCGG ......C...
MS2             ..A....... .A.....C.. C.....C... T.G.CGCCGG ......A...
M12             ..A....... .A.....C.. C.....C... T.G.TACCGG ......A...
```

fr ..C..... .C.....G.. A.....C... G.A.CCCGTG ......C...

```
                        ORF4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1760       1770       1780       1790       1800
Consensus       ATGAGGAWTA CCCATGTCRA ARWCAACAAA GAAGTTCAAC TCTTTATGTA
DL1             .......T.. ........G. .GA....... .......... ..........
DL2             .......T.. ........G. .GA....... .......... ..........
DL13            .......T.. ........G. .GA....... .......... ..........
DL16            .......T.. ........G. .GA....... .......... ..........
ST4             .......T.. ........G. .GA....... .......... ..........
R17             .......T.. ........G. .GA....... .......... ..........
J20             .......T.. ........G. .GA....... .......... ..........
MS2             .......T.. ........G. .GA....... .......... ..........
M12             .......T.. ........G. ..A....... .......... ..........
fr              .......A.. ........A. .AT....... .......... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1810       1820       1830       1840       1850
Consensus       TTGATYTKYC TYGCGATCTT TCTCTCGAAR TTTACCAATC AATTGCTWSY
DL1             .....C.TC. .C........ .........A .......... .......TCT
DL2             .....C.TC. .C........ .........A .......... .......TCT
DL13            .....C.TC. .C........ .........A .......... .......TCT
DL16            .....C.TC. .T........ .........A .......... .......TCT
ST4             .....C.TC. .C........ .........A .......... .......TCT
R17             .....C.TC. .C........ .........A .......... .......TCT
J20             .....C.TC. .C........ .........A .......... .......TCT
MS2             .....C.TC. .C........ .........A .......... .......TCT
M12             .....C.TC. .C........ .........A .......... .......TCT
fr              .....T.GT. .C........ .........G .......... .......AGC

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1860       1870       1880       1890       1900
Consensus       GTCGCTACTG GAWSYKSTRA TCCGCAYAGT RRMGACTTTW CRGCAATTGC
DL1             .......... ..AGCGG.G. ......C... AGA......A .A........
DL2             .......... ..AGCGG.G. ......C... AGA......A .A........
DL13            .......... ..AGCGG.G. ......C... AGA......A .A........
DL16            .......... ..AGCGG.G. ......C... AGA......A .A........
ST4             .......... ..AGCGG.G. ......C... GAC......A .A........
R17             .......... ..AGCGG.A. ......C... GAC......A .A........
J20             .......... ..AGCGG.G. ......C... AGA......A .G........
MS2             .......... ..AGCGG.G. ......C... GAC......A .A........
M12             .......... ..TGCGG.A. ......C... GAC......T .A........
fr              .......... ..TCTTC.G. ......T... GAC......A .A........
                     STOP 3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1910       1920       1930       1940       1950
Consensus       TTACYTAAGR GACGARYTKY THACWAAGCA YCCSWMBYTA GGNDMYGGTA
DL1             ....C....G .....GT.GC .A..T..... T..CTCCT.. ..CAAT....
DL2             ....C....A .....AT.GC .A..T..... T..GTCCT.. ..AAAT....
DL13            ....C....A .....AT.GC .A..T..... T..GTCCT.. ..AAAT....
DL16            ....C....A .....AT.GC .A..T..... T..GTCCT.. ..AAAT....
ST4             ....T....G .....AT.GC .C..A..... T..GACCT.. ..TTCT....
R17             ....C....G .....AT.GC .T..A..... T..GACTT.. ..CTCT....
J20             ....C....A .....GT.GC .C..A..... T..GTCTT.. ..AAAT....
MS2             ....T....G .....AT.GC .C..A..... T..GACCT.. ..TTCT....
M12             ....C....G .....GT.GC .T..T..... C..GACGT.. ..TTCC....
```

fr ....C....G .....GC.TT .A..A..... T..GAACC.. ..GGAT....

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1960       1970       1980       1990       2000
Consensus       AYGACGARGC RACCCGYMGN RSYYTAGCTA TYGCTAAGCT NCKGGAGGCG
DL1             .T.....G.. G.....CC.A GCTT...... .T........ T.G.......
DL2             .T.....G.. G.....CC.G GCTC...... .C........ C.G.......
DL13            .T.....G.. G.....CC.G GCTC...... .C........ C.G.......
DL16            .T.....G.. G.....CC.G GCTC...... .C........ C.G.......
ST4             .T.....G.. G.....TC.T ACCT...... .C........ A.G.......
R17             .T.....G.. G.....TC.C ACCT...... .T........ G.G.......
J20             .T.....G.. G.....CC.G GCTT...... .C........ C.G.......
MS2             .T.....G.. G.....TC.T ACCT...... .C........ A.G.......
M12             .T.....G.. G.....CC.T GCTT...... .T........ A.G.......
fr              .C.....A.. A.....CA.A AGCC...... .C........ G.T.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2010       2020       2030       2040       2050
Consensus       AATGRNSRDY GYGGYCAGAT HAAYAGRGAD GGTTTCTTAC AYGAYRMAWC
DL1             ....AACGAT .T..C..... C..T..A..G .......... .T..CAA.T.
DL2             ....AGCGGT .C..C..... T..T..G..A .......... .C..TAA.T.
DL13            ....AGCGGT .C..C..... T..T..G..A .......... .C..TAA.T.
DL16            ....AGCGGT .C..C..... T..T..G..A .......... .C..TAA.T.
ST4             ....ATCGGT .C..T..... A..T..A..A .......... .T..CAA.T.
R17             ....ATCGGT .C..C..... A..T..A..A .......... .T..CAA.T.
J20             ....AACGAT .T..C..... T..C..A..A .......... .T..CAA.T.
MS2             ....GTGATC .C..T..... A..T..A..A .......... .T..CAA.T.
M12             ....ATCGAT .T..C..... T..C..A..A .......... .T..CAA.T.
fr              ....ACCGTT .T..T..... A..T..G..T .......... .T..CGC.A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2060       2070       2080       2090       2100
Consensus       CKYRTCRTGG GATCCGGATG TTTTACAAAC CAGCATCCGT AGCCTWATHG
DL1             .TTG..A... .......... .......... .......... .....T..T.
DL2             .TTA..G... .......... .......... .......... .....T..T.
DL13            .TTA..G... .......... .......... .......... .....T..T.
DL16            .TTA..G... .......... .......... .......... .....T..T.
ST4             .TTG..A... .......... .......... .......... .....T..T.
R17             .TTG..A... .......... .......... .......... .....T..C.
J20             .TTG..G... .......... .......... .......... .....T..T.
MS2             .TTG..A... .......... .......... .......... .....T..T.
M12             .TTG..G... .......... .......... .......... .....A..A.
fr              .GCG..G... .......... .......... .......... .....T..T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2110       2120       2130       2140       2150
Consensus       GYAAYCTYCT YTCTGGYTAY MRHWSKYMGT TGTTTRGRCA MTGYACRTTY
DL1             .C..C..T.. C.....C..T CGTTCGTC.. .....G.G.. A..C..G..C
DL2             .C..C..T.. C.....C..C CGCTCGTC.. .....G.G.. A..C..G..C
DL13            .C..C..T.. C.....C..C CGCTCGTC.. .....G.G.. A..C..G..C
DL16            .C..C..T.. C.....C..C CGCTCGTC.. .....G.G.. A..C..G..C
ST4             .C..C..C.. C.....C..C CGATCGTC.. .....G.G.. A..C..G..C
R17             .C..T..T.. C.....C..C CAATCGTC.. .....G.G.. A..C..G..C
J20             .C..C..T.. C.....C..T CGTTCGTC.. .....G.G.. A..C..G..T
MS2             .C..C..C.. C.....C..C CGATCGTC.. .....G.G.. A..C..G..C
```

```
M12          .C..C..C.. C.....T..C CGATCGTC.. .....G.G.. A..C..G..C
fr           .T..C..C.. T.....C..T AGCAGTCA.. .....A.A.. C..T..A..C
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2160       2170       2180       2190       2200
Consensus       TCMAACGGTG CYYCDATGGG GCACAAGTTG CAGGATGCAG CGCCTTAYAA
DL1             ..C....... .CT.G..... .......... .......... .......C..
DL2             ..C....... .CT.A..... .......... .......... .......C..
DL13            ..C....... .CT.A..... .......... .......... .......C..
DL16            ..C....... .CT.A..... .......... .......... .......C..
ST4             ..C....... .CT.T..... .......... .......... .......C..
R17             ..C....... .CT.T..... .......... .......... .......C..
J20             ..C....... .CT.T..... .......... .......... .......C..
MS2             ..C....... .TC.T..... .......... .......... .......C..
M12             ..C....... .CT.G..... .......... .......... .......C..
fr              ..A....... .CT.T..... .......... .......... .......T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2210       2220       2230       2240       2250
Consensus       GAAGTTCGCT GAACAAGCAA CCGTKACSCC SMGSGCKYTR ARAGCGGCNB
DL1             .......... .......... ....T..C.. CC.C..TC.A .G......CC
DL2             .......... .......... ....T..C.. CC.C..TC.G .G......AC
DL13            .......... .......... ....T..C.. CC.C..TC.G .G......AC
DL16            .......... .......... ....T..C.. CC.C..TC.G .G......AC
ST4             .......... .......... ....T..C.. CC.C..TC.G .G......TC
R17             .......... .......... ....T..C.. CC.C..TC.G .G......TC
J20             .......... .......... ....T..C.. CC.C..TC.G .G......GT
MS2             .......... .......... ....T..C.. CC.C..TC.G .G......TC
M12             .......... .......... ....T..C.. CC.C..TC.A .G......CC
fr              .......... .......... ....G..G.. GA.G..GT.G .A......AG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2260       2270       2280       2290       2300
Consensus       TRYTGGTCMR AGAYCARTGY RBKCCSTGGA TYMGWCACKC GSWCSDCTWY
DL1             .AC.....CG ...C..G..T GTG..G.... .TA.A...G. .GT.CG..AC
DL2             .AC.....CG ...C..G..T GTG..G.... .CA.A...G. .GT.CG..AC
DL13            .AC.....CG ...C..G..T GTG..G.... .CA.A...G. .GT.CG..AC
DL16            .AC.....CG ...C..G..T GTG..G.... .CA.A...G. .GT.CG..AC
ST4             .AT.....CG ...C..A..T GCG..G.... .CA.A...G. .GT.CG..AT
R17             .AT.....CG ...C..A..T GCG..G.... .CA.A...G. .GT.CA..AT
J20             .AC.....AG ...C..G..T GCG..G.... .CA.A...G. .GT.CG..AT
MS2             .AT.....CG ...C..A..T GCG..G.... .CA.A...G. .GT.CG..AT
M12             .GC.....AG ...C..G..T GCG..G.... .TA.A...G. .GT.CG..AC
fr              .GC.....AA ...T..G..C AGT..C.... .CC.T...T. .CA.GT..TC


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2310       2320       2330       2340       2350
Consensus       MMYGARTCAT ATRMATTTAG RCTCGTBGKA GGSAACGGWG TGTTYACWGT
DL1             AAT..G.... ..GA...... G.....T.T. ..G.....A. ....T..A..
DL2             AAC..G.... ..GA...... G.....C.T. ..G.....A. ....T..A..
DL13            AAC..G.... ..GA...... G.....C.T. ..G.....A. ....T..A..
DL16            AAC..G.... ..GA...... G.....C.T. ..G.....A. ....T..A..
ST4             AAC..G.... ..GA...... G.....T.T. ..G.....A. ....T..A..
R17             AAC..G.... ..GA...... G.....T.T. ..G.....A. ....T..A..
J20             AAC..A.... ..AA...... A.....T.T. ..G.....A. ....T..A..
MS2             AAC..G.... ..GA...... G.....T.T. ..G.....A. ....T..A..
```

```
M12          AAC..G.... ..GA...... G.....T.T. ..G.....A. ....C..A..
fr           CCC..G.... ..AC...... G.....G.G. ..C.....T. ....T..T..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2360       2370       2380       2390       2400
Consensus       TCCGAAGAAY AATAAAATAG ATCGGGCTGC YTGYAARGAG CCYGATATGA
DL1             .........T .......... .......... C..T..G... ..C.......
DL2             .........T .......... .......... C..T..G... ..C.......
DL13            .........T .......... .......... C..T..G... ..C.......
DL16            .........T .......... .......... C..T..G... ..C.......
ST4             .........T .......... .......... C..T..G... ..T.......
R17             .........T .......... .......... C..T..G... ..T.......
J20             .........T .......... .......... C..T..G... ..C.......
MS2             .........T .......... .......... C..T..G... ..T.......
M12             .........T .......... .......... T..C..A... ..C.......
fr              .........C .......... .......... T..C..A... ..T.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2410       2420       2430       2440       2450
Consensus       ATATGTACYT HCAGAAAGGR GTHGGHGSYT TYATHMGDCG YCGBCTYARR
DL1             ........C. T........G ..T..C.CC. .C..TA.A.. C..C..C.AA
DL2             ........C. C........A ..T..C.CC. .C..AA.G.. T..C..C.AA
DL13            ........C. C........A ..T..C.CC. .C..AA.G.. T..C..C.AA
DL16            ........C. C........A ..T..C.CC. .C..AA.G.. T..C..C.AA
ST4             ........C. C........G ..C..T.CC. .T..CA.A.. C..G..C.AA
R17             ........C. C........G ..C..C.CT. .T..TA.A.. C..G..C.AA
J20             ........C. T........A ..T..C.CT. .T..AC.G.. T..T..C.GA
MS2             ........C. C........G ..C..T.CT. .C..CA.A.. C..G..C.AA
M12             ........C. T........A ..A..A.CC. .C..CA.G.. C..T..C.AA
fr              ........T. A........G ..C..C.GT. .C..CC.T.. C..C..T.AG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2460       2470       2480       2490       2500
Consensus       WCYGTBGGTA TAGAYYTGAA YGATCARWCG ATCAAYCARC KYYTRGCTCA
DL1             T.T..C.... ....CC.... T.....AA.. .....T..G. GTC.A.....
DL2             T.T..C.... ....CC.... T.....AA.. .....T..G. GTT.A.....
DL13            T.T..C.... ....CC.... T.....AA.. .....T..G. GTT.A.....
DL16            T.T..C.... ....CC.... T.....AA.. .....T..G. GTT.A.....
ST4             T.C..T.... ....TC.... T.....AT.. .....C..G. TTC.G.....
R17             T.C..T.... ....CC.... T.....GT.. .....C..G. GTC.A.....
J20             T.C..T.... ....CC.... T.....AA.. .....C..G. GTC.G.....
MS2             T.C..T.... ....CC.... T.....AT.. .....C..G. GTC.G.....
M12             T.C..T.... ....CC.... T.....AA.. .....C..G. GTC.G.....
fr              A.T..G.... ....CT.... C.....AA.. .....T..A. GCC.G.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2510       2520       2530       2540       2550
Consensus       RCARGGYAGY VBHGATGGNT CDYTDGCRAC KATAGAYYTA TCGTCNGCNT
DL1             A..G..C..C ATA.....T. .TT.A..G.. G.....CC.. .....A..T.
DL2             A..G..C..C ATA.....C. .TT.A..A.. G.....TT.. .....G..T.
DL13            A..G..C..C ATA.....C. .TT.A..A.. G.....TT.. .....G..T.
DL16            A..G..C..C ATA.....C. .TT.A..A.. G.....TT.. .....G..T.
ST4             G..G..C..C GTA.....T. .GC.T..G.. G.....CT.. .....T..A.
R17             G..A..C..C GCA.....T. .GC.T..G.. G.....CT.. .....C..G.
J20             A..A..C..C GTC.....C. .AT.G..G.. G.....CT.. .....G..C.
MS2             G..G..C..C GTA.....T. .GC.T..G.. G.....CT.. .....T..A.
```

```
M12          A..A..C..T ATA.....A. .GT.A..G.. T.....CT.. .....T..A.
fr           A..A..T..C CGT.....G. .TT.G..G.. G.....CT.. .....T..T.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2560       2570       2580       2590       2600
Consensus       CBGAYTCYAT YWSYGAYCGC CTRGTGTGGA RYTTYCTYCC ACCTGAGYTA
DL1             .G..T..C.. CTCT..C... ..G.....  GC..T..C.. .......C..
DL2             .T..T..T.. CTCT..C... ..G.....  GT..C..C.. .......T..
DL13            .T..T..T.. CTCT..C... ..G.....  GT..C..C.. .......T..
DL16            .T..T..T.. CTCT..C... ..G.....  GT..C..C.. .......T..
ST4             .C..T..C.. CTCC..T... ..G.....  GT..T..C.. .......C..
R17             .C..T..C.. TTCC..C... ..A.....  AT..C..T.. .......C..
J20             .T..C..C.. CTCT..C... ..G.....  GT..C..C.. .......C..
MS2             .C..T..C.. CTCC..T... ..G.....  GT..T..C.. .......C..
M12             .C..T..C.. CTCT..C... ..G.....  GT..C..C.. .......C..
fr              .T..C..C.. CAGC..C... ..A.....  GT..T..C.. .......C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2610       2620       2630       2640       2650
Consensus       TAYTCATATC TYGAYMKKAT YCGMWSVCAC TAYGGWWWCR TARATGGNRA
DL1             ..T.......  .C..TCGT.. C..CTCC... ..C..AAT.G ..G....CG.
DL2             ..T.......  .C..TCGT.. C..CTCC... ..C..AAT.G ..G....GG.
DL13            ..T.......  .C..TCGT.. C..CTCC... ..C..AAT.G ..G....GG.
DL16            ..T.......  .C..TCGT.. C..CTCC... ..C..AAT.G ..G....GG.
ST4             ..T.......  .C..TCGT.. C..CTCA... ..C..AAT.G ..G....CG.
R17             ..T.......  .T..TCGT.. C..CTCG... ..C..AAT.G ..G....CG.
J20             ..T.......  .C..TCGT.. C..CTCC... ..T..AAT.A ..G....AG.
MS2             ..T.......  .C..TCGT.. C..CTCA... ..C..AAT.G ..G....CG.
M12             ..C.......  .C..TCGT.. C..CTCC... ..C..AAT.G ..G....TG.
fr              ..T.......  .C..CATG.. T..AAGC... ..C..TTA.G ..A....CA.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2660       2670       2680       2690       2700
Consensus       GAYGATWCGD TGGGAACTAT TTTCSACRAT GGGWAAYGGG TTYACNTTYG
DL1             ..C...A..G .......... ....C..A.. ...A..T... ..C..T..C.
DL2             ..C...A..A .......... ....C..A.. ...T..T... ..C..T..C.
DL13            ..C...A..A .......... ....C..A.. ...T..T... ..C..T..C.
DL16            ..C...A..A .......... ....C..A.. ...T..T... ..C..T..C.
ST4             ..C...A..A .......... ....C..A.. ...A..T... ..C..G..T.
R17             ..C...A..A .......... ....C..G.. ...A..T... ..T..A..T.
J20             ..C...A..A .......... ....C..A.. ...A..T... ..C..T..T.
MS2             ..C...A..A .......... ....C..A.. ...A..T... ..C..A..T.
M12             ..C...A..A .......... ....C..A.. ...A..C... ..C..T..T.
fr              ..T...T..T .......... ....G..G.. ...T..T... ..C..C..T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2710       2720       2730       2740       2750
Consensus       ARCTAGAGTC CATGATHTTY TGGGCWATAG TVARRGCDAC YCARATCCAT
DL1             .A........ ......A..T .....T.... .G.AA..G.. C..G......
DL2             .A........ ......C..T .....T.... .G.AA..A.. C..G......
DL13            .A........ ......C..T .....T.... .G.AA..A.. C..G......
DL16            .A........ ......C..T .....T.... .G.AA..A.. C..G......
ST4             .G........ ......A..C .....A.... .C.AA..G.. C..A......
R17             .G........ ......A..C .....A.... .C.AA..A.. C..A......
J20             .A........ ......A..T .....T.... .A.AA..A.. C..G......
MS2             .G........ ......A..C .....A.... .C.AA..G.. C..A......
```

```
M12          .G........ ......A..T .....T.... .C.AA..G.. C..A......
fr           .A........ ......T..C .....T.... .C.GG..T.. T..G......
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2760       2770       2780       2790       2800
Consensus       TTTSGTAACR CCGGAACCAT WGGCATCTAY GGGGACGATA TYATATGYCC
DL1             ...G.....G .......... A........C .......... .T.....T..
DL2             ...G.....G .......... A........C .......... .T.....T..
DL13            ...G.....G .......... A........C .......... .T.....T..
DL16            ...G.....G .......... A........C .......... .T.....T..
ST4             ...G.....G .......... A........C .......... .T.....T..
R17             ...G.....G .......... A........C .......... .C.....T..
J20             ...G.....G .......... A........C .......... .T.....T..
MS2             ...G.....G .......... A........C .......... .T.....T..
M12             ...G.....G .......... A........C .......... .T.....T..
fr              ...C.....A .......... T........T .......... .T.....C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2810       2820       2830       2840       2850
Consensus       CASWGAGATT GCACCYCGYG TGCTRGARGC RCTDRSCTWC TACGGTTTYA
DL1             ..GT...... .....C..T. ....A..G.. A..TGC..A. ........T.
DL2             ..GT...... .....C..T. ....A..G.. A..TGC..A. ........C.
DL13            ..GT...... .....C..T. ....A..G.. A..TGC..A. ........C.
DL16            ..GT...... .....C..T. ....A..G.. A..TGC..A. ........C.
ST4             ..GT...... .....C..T. ....G..G.. A..TGC..A. ........C.
R17             ..GT...... .....C..T. ....A..G.. A..TGC..A. ........T.
J20             ..GT...... .....C..T. ....A..G.. G..AGC..A. ........C.
MS2             ..GT...... .....C..T. ....A..G.. A..TGC..A. ........T.
M12             ..GT...... .....C..T. ....A..G.. A..TGC..A. ........T.
fr              ..CA...... .....T..C. ....G..A.. A..GAG..T. ........C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2860       2870       2880       2890       2900
Consensus       AACCGAATCT HCGWAARACG TTCRYGTCVG GSYYYTTTCG CGAGAGCTGY
DL1             .......... T..T..A... ...GT...G. .GCTC..... .........C
DL2             .......... T..T..A... ...GT...A. .GCTC..... .........T
DL13            .......... T..T..A... ...GT...A. .GCTC..... .........T
DL16            .......... T..T..A... ...GT...A. .GCTC..... .........T
ST4             .......... C..T..A... ...GT...C. .GCTC..... .........C
R17             .......... T..T..A... ...GT...C. .GCTC..... .........C
J20             .......... T..T..A... ...GT...A. .GCTC..... .........T
MS2             .......... T..T..A... ...GT...C. .GCTC..... .........C
M12             .......... T..T..A... ...GT...C. .GCTC..... .........C
fr              .......... A..A..G... ...AC...C. .CTCT..... .........C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2910       2920       2930       2940       2950
Consensus       RGCGCGCACT WTTWCCGTGG TGTCGATGTY AAACCRTTYT AYATCARGAA
DL1             G......... T..A...... .........C .....G..T. .C....A...
DL2             G......... T..A...... .........C .....G..C. .C....A...
DL13            G......... T..A...... .........C .....G..C. .C....A...
DL16            G......... T..A...... .........C .....G..C. .C....A...
ST4             G......... T..A...... .........C .....G..T. .C....A...
R17             A......... T..A...... .........C .....G..T. .C....A...
J20             G......... T..A...... .........C .....G..T. .C....G...
MS2             G......... T..A...... .........C .....G..T. .C....A...
```

```
M12          G......... T..A...... .........C .....G..C. .C....A...
fr           G......... A..T...... .........T .....A..T. .T....A...
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2960       2970       2980       2990       3000
Consensus       ACCWRTYRMY RAYCTMTTYK CCCTDATGCT KATMHTDAAY CGKMTWMGSG
DL1             ...AG.TGAC A.C..C..CT ....T..... G..AA.G..T ..GC.TA.G.
DL2             ...AG.TGAC A.T..C..CT ....T..... G..CA.G..T ..GC.TA.G.
DL13            ...AG.TGAC A.T..C..CT ....T..... G..CA.G..T ..GC.TA.G.
DL16            ...AG.TGAC A.T..C..CT ....T..... G..CA.G..T ..GC.TA.G.
ST4             ...TG.TGAC A.T..C..CG ....T..... G..AT.G..T ..GC.AC.G.
R17             ...TG.TGAC A.C..C..TG ....G..... G..AT.A..T ..GC.AC.G.
J20             ...AG.CGAC A.T..C..CT ....T..... G..CA.G..T ..GC.TA.G.
MS2             ...TG.TGAC A.T..C..CG ....G..... G..AT.A..T ..GC.AC.G.
M12             ...AG.TGAC A.C..C..TT ....T..... G..AC.G..T ..GC.AC.G.
fr              ...AA.CACT G.C..A..CT ....A..... T..AC.T..C ..TA.AC.C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3010       3020       3030       3040       3050
Consensus       GDTGGGGNGT WGTCRRMGGW ATRKCAGAYC CDCGCCTHTA YRAGGTDTGG
DL1             .G.....C.. T...GGA..T ..GT....T. .G.....T.. TA....T...
DL2             .G.....C.. T...GGA..T ..GT....T. .G.....T.. TA....T...
DL13            .G.....C.. T...GGA..T ..GT....T. .G.....T.. TA....T...
DL16            .G.....C.. T...GGA..T ..GT....T. .G.....T.. TA....T...
ST4             .T.....A.. T...GGA..T ..GT....T. .A.....T.. CA....G...
R17             .T.....G.. T...GGA..T ..GT....T. .A.....T.. CA....G...
J20             .A.....T.. T...GGA..T ..GT....T. .T.....A.. TA....T...
MS2             .T.....A.. T...GGA..T ..GT....T. .A.....C.. TA....G...
M12             .T.....G.. T...GGA..T ..GT....T. .G.....T.. CA....T...
fr              .A.....G.. A...AAC..A ..AG....C. .A.....C.. CG....A...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3060       3070       3080       3090       3100
Consensus       GWRMRRCTVT CCWSMYWGGT DCCDWSRWWK TTHTTYGGTS SSACGGACCT
DL1             .TACGA..G. ..TCCCT... T..ATCGATG ..C..C...G GG........
DL2             .TACGA..A. ..TCCCT... T..ATCGATG ..C..C...G GG........
DL13            .TACGA..A. ..TCCCT... T..ATCGATG ..C..C...G GG........
DL16            .TACGA..A. ..TCCCT... T..ATCGATG ..C..C...G GG........
ST4             .TACGA..C. ..TCCCA... G..TTCGATG ..T..C...G GG........
R17             .TACGA..C. ..TCCCA... A..TTCGATG ..C..C...G GG........
J20             .TACGA..G. ..TCCCT... T..ATCGATG ..C..C...G GG........
MS2             .TACGG..C. ..TCCCA... G..TTCGATG ..C..C...G GG........
M12             .TACGA..G. ..TCCCT... T..ATCGATG ..C..C...C CC........
fr              .AGAAG..C. ..AGATT... A..GAGATAT ..A..T...G GG........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3110       3120       3130       3140       3150
Consensus       BSMKGCYGAY TAYTACGTMG TBWSMCCBCC NAHNSYDRWM KSRRTWTAYW
DL1             TGCT..C..C ..T.....A. .CAGC..C.. G.CCGCAGTC TCAG.T..TA
DL2             TGCT..C..C ..C.....A. .TAGC..T.. G.ATGCTGTC TCGG.T..TA
DL13            TGCT..C..C ..C.....A. .TAGC..T.. G.ATGCTGTC TCGG.T..TA
DL16            TGCT..C..C ..C.....A. .TAGC..T.. G.ATGCTGTC TCGG.T..TA
ST4             CGCT..C..C ..C.....A. .CAGC..G.. C.CGGCAGTC TCGG.A..TA
R17             CGCT..C..C ..C.....A. .CAGC..G.. T.CGGCGGTC TCGG.A..TA
J20             TGCT..T..C ..C.....A. .CAGC..C.. G.CAGCGGTC TCAG.T..TA
MS2             CGCT..C..C ..C.....A. .CAGC..G.. T.CGGCAGTC TCGG.A..CA
```

```
M12          CGCT..T..C ..C.....A. .CAGC..T.. A.CTGCTGTC TCAG.A..TA
fr           GCAG..C..T ..C.....C. .GTCA..T.. T.TTCTGAAA GGAA.T..TT
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3160       3170       3180       3190       3200
Consensus       CYAARACYSM VTRBGRRMSG YKRSTAYGCS GAHRCYMGWA CMWCRGGYTT
DL1             .T..A..TGC C.AC.GGAG. TTAC.-C..G ..TA.CC.T. .CT.G..T..
DL2             .C..A..TGC A.AT.GGAG. TTAC.-C..G ..CG.CC.T. .CT.G..T..
DL13            .C..A..TGC A.AT.GGAG. TTAC.-C..G ..CG.CC.T. .CT.G..T..
DL16            .C..A..TGC A.AT.GGAG. TTAC.-C..G ..CG.CC.T. .CT.G..T..
ST4             .C..G..TCC G.AT.GGCG. CTAC.-C..G ..TA.CC.T. .CT.G..T..
R17             .C..G..TCC G.AT.GACG. CTGC.-C..G ..TA.CC.T. .CT.G..T..
J20             .T..G..CGC A.AT.GGAG. TTGC.-C..G ..CG.CC.T. .CT.G..T..
MS2             .C..G..TCC G.AC.GGCG. CTGC.-C..G ..TA.CC.T. .CT.G..T..
M12             .T..G..CGC G.AT.GGAG. CTGC.-C..G ..TA.CC.T. .CT.G..T..
fr              .C..G.-TGA A.GG.AGAC. CGAG..T..C ..AG.TA.A. .AA.A..C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3210       3220       3230       3240       3250
Consensus       CMRKCTYGCT MGHAYCGCDM RRKRRCGHAA GYDCTTYWSC GADAAGCAYG
DL1             .CGT..T... C.T.T...AA AAGAG..A.. .CG...TAG. ..G.....T.
DL2             .CGT..T... C.T.T...GA AAGAG..A.. .CA...TAG. ..G.....T.
DL13            .CGT..T... C.T.T...GA AAGAG..A.. .CA...TAG. ..G.....T.
DL16            .CGT..T... C.T.C...GA AAGAG..A.. .CA...TAG. ..G.....T.
ST4             .CGT..T... C.T.T...TC GAGAA..C.. .TT...CAG. ..A.....T.
R17             .CGT..T... C.T.T...TC GAGAA..C.. .TT...CAG. ..A.....T.
J20             .CGT..T... C.T.T...GA AAGAG..T.. .CA...TAG. ..G.....T.
MS2             .CGT..T... C.T.T...TC GAGAA..C.. .TT...CAG. ..A.....C.
M12             .CGT..T... A.A.T...AA AAGAG..T.. .CG...TAG. ..G.....T.
fr              .AAG..C... C.C.T...GA GGTGG..A.. .CA...TTC. ..T.....C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3260       3270       3280       3290       3300
Consensus       RCWSWGGYCG STACATWGCR TGGTTCCATA CTGGAGGTGA RRTCACCGAY
DL1             A.AGT..T.. C.....A..A .......... .......... GA.......T
DL2             A.AGT..T.. C.....A..A .......... .......... AA.......T
DL13            G.AGT..T.. C.....A..A .......... .......... AA.......T
DL16            A.AGT..T.. C.....A..A .......... .......... AA.......T
ST4             A.AGT..C.. C.....A..G .......... .......... AG.......C
R17             A.AGT..T.. C.....A..G .......... .......... AA.......C
J20             A.AGT..T.. C.....A..A .......... .......... AA.......T
MS2             A.AGT..T.. C.....A..G .......... .......... AA.......C
M12             A.AGT..T.. C.....A..G .......... .......... AA.......T
fr              A.TCA..T.. G.....T..G .......... .......... GA.......T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3310       3320       3330       3340       3350
Consensus       AGYATGAAGT CCGCYGGYGT RCGYRTHMTR CGYACTTCGG ARTGGCTRMM
DL1             ..T....... ....T..C.. G..CG.AA.G ..C....... .G.....GAC
DL2             ..T....... ....T..C.. G..TG.AA.A ..C....... .G.....GAC
DL13            ..T....... ....T..C.. G..TG.AA.A ..C....... .G.....GAC
DL16            ..T....... ....T..C.. G..TG.AA.A ..C....... .G.....GAC
ST4             ..T....... ....C..C.. G..TA.TA.G ..C....... .G.....AAC
R17             ..T....... ....C..C.. G..CA.CA.G ..C....... .G.....AAC
J20             ..T....... ....C..C.. G..TG.TA.G ..C....... .A.....AAC
MS2             ..C....... ....C..C.. G..CG.TA.A ..C....... .G.....AAC
```

```
M12          ..T....... ....T..C.. G..CG.AC.G ..C....... .G..------
fr           ..T....... ....C..T.. A..CG.TA.G ..T....... .G.....ACA
```

```
                                                                                                                STOP 4
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        3360       3370       3380       3390       3400
Consensus       GCCGGTTCCC RYRTTCCCKC AGGARTGYGG GCCAGCGAGC TCTCCTCRGT
DL1             .......... ACA.....T. ....G..T.. .......... .......G..
DL2             .......... ATA.....T. ....A..T.. .......... .......G..
DL13            .......... ATA.....T. ....A..T.. .......... .......G..
DL16            .......... ATA.....T. ....A..T.. .......... .......G..
ST4             .......... ACA.....T. ....G..T.. .......... .......G..
R17             .......... ACA.....T. ....G..T.. .......... .......G..
J20             .......... ACA.....G. ....G..T.. .......... .......G..
MS2             .......... ACA.....T. ....G..T.. .......... .......G..
M12             ---------- ---------- ---------- ---------- ----------
fr              .......... GTG.....T. ....A..C.. .......... .......A..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3410       3420       3430       3440       3450
Consensus       AGCWRAACYG AGGGACCCCC GTAARCGGGG TGGGTGTGCT CGAAAGAGCA
DL1             ...TG.-.C. .......... ....A..... .......... ..........
DL2             ...TG.-.C. .......... ....A..... .......... ..........
DL13            ...TG.-.C. .......... ....A..... .......... ..........
DL16            ...TG.-.C. .......... ....A..... .......... ..........
ST4             ...TG.-.C. .......... ....A..... .......... ..........
R17             ...TG.-.C. .......... ....A..... .......... ..........
J20             ...TG.-.C. .......... ....A..... .......... ..........
MS2             ...TG.-.C. .......... ....A..... .......... ..........
M12             ---------- ---------- ---------- ---------- ----------
fr              ...AA...T. .......... ....G..... .......... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3460       3470       3480       3490       3500
Consensus       CGGGTCCGCG WAAGCGGTCC GGCTCMWCMG AAATGGTGGG YGAGGSTTCG
DL1             .......... A.......-- .....CA.C. ...-...... C.-..C....
DL2             .......... A.......-- .....CA.C. ...-...... C.-..C....
DL13            .......... A.......-- .....CA.C. ...-...... C.-..C....
DL16            .......... A.......-- .....CA.C. ...-...... C.-..C....
ST4             .......... A.......-- .....CA.C. ...-...... C.-..C....
R17             .......... A.......-- .....CA.C. ...-...... C.-..C....
J20             .......... T.......-- .....CA.C. ...-...... C.-..C....
MS2             .....--... A......... .....CA.C. ...-...... C.-..C....
M12             ---------- ---------- ---------- ---------- ----------
fr              .......... A.......-- .....AT.A. .......... T....G....

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3510       3520       3530       3540       3550
Consensus       SCCTCWRGGA CYKSCCCYKR AAGRKMRGRC CCGGGATTCT CCCGATTTTG
DL1             G..-.AG... .CTC...CTA ...AGAG.A. .......... ........-.
DL2             G..-.AG... .CTC...CTA ...AGAG.A. .......... ........-.
DL13            G..-.AG... .CTC...CTA ...AGAG.A. .......... ........-.
DL16            G..-.AG... .CTC...CTA ...AGAG.A. .......... ........-.
ST4             G..-.AG... .CTC...TTG ...AGAG.G. .......... ........-.
R17             G..-.AG... .CTC...CTG ...AGAG.A. .......... ........-.
J20             G..-.AG... .CTC...CTA ...AGAG.A. .......... ........-.
MS2             G..-.AG... .CTC...CTA ...AGAG.A. .......... ........-.
M12             ---------- ---------- ---------- ---------- ----------
```

```
fr          C....TA... .TGG...CGA ...GTCA.G. .......... ..........
```

```
                ....|....| ....|....| ....|....|
                        3560       3570       3580
Consensus       GTARCTAGYT GCTTGGCTAG YKACCACCCA
DL1             ...A....C. .......... TT........
DL2             ...A....C. .......... TT........
DL13            ...A....C. .......... TT........
DL16            ...A....C. .......... TT........
ST4             ...A....C. .......... TT........
R17             ...A....C. .......... TG........
J20             ...A....C. .......... TT........
MS2             ...A....C. .......... TT........
M12             ---------- ---------- ----------
fr              ...G....T. .......... CT........
```

Appendix A2
Group II

Alignment: Levivirus Group II.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       GGKTGGBKVS SMHYYYDKKS SGSKTBCYBY YHHMCTTSMW GKCGMKMKRR
T72             ---------- ---------- ---------- ---------- ----------
DL10            ---------- ---------- ---------- ---------- ----------
DL20            -----..GTGG GACCCCTTTC G.GG.C.TGC TCAA...CCT .T..AGCTAA
GA              -.G...TTAG GAATTTAGGG G.GG.T.CCT CACC...GAA .G..CTAGGA
KU1             ..T...CGCC CCTTTCGGGG C.CT.G.TTT CTTC...GAA .G..CTAGGG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       WKSBWTTKYW WWARYGDCTY WWKYSWGMMG TCKTTYYAYC ATGCCGTTAG
T72             ---------- ---------- ---------- ---------- ..........
DL10            ---------- ---------- ---------- ---------- ..........
DL20            TGCCA..TTT TT.AT.T..T TAGCGA.AC. -----CT.C. ..........
GA              ATGGT..GCA TA.GC.G..T AATTCT.CA. ..G..CC.C. ..........
KU1             ATGTT..GCA AA.GC.A..C ATGCCT.CA. ..T..TC.T. ..........
                                                      ORF1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       GTTTAGRTGA CGGTATRTTC CAYATACCGG AGGADCTATG TTTCCGAARY
T72             ......G... ......A.-. ..T....... ....G..... ........AC
DL10            ......A... ......A.-. ..T....... ....T..... ........GT
DL20            ......A... ......A... ..T....... ....T..... ........GT
GA              ......G... ......G.-. ..C....... ....T..... ........GT
KU1             ......A... ......A.-. ..T....... ....A..... ........AC

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       MRAATATMGA YMGAAHYTAY MAKGTTAMAC TTRTHTCTTA CGAYRAKAAR
T72             AG.....C.. TC...CC..C A.G....C.. ..A.T..... ...CA.G..G
DL10            CG.....A.. CA...AT..C A.G....C.. ..A.A..... ...CG.G..A
DL20            CA.....C.. CA...AT..C A.G....A.. ..A.T..... ...CA.G..G
GA              CA.....A.. CA...AT..C A.G....A.. ..A.A..... ...CA.G..A
KU1             AG.....C.. TC...TT..T C.T....A.. ..G.C..... ...TA.T..A

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       GGDAADSTHR YTTCYGAYGA YTCTTTYGAR YMDRYMGARA AYTAYCTCYT
T72             ..T..TG.CA C...T..T.. T.....C..A TCGACA..G. .C..T...C.
DL10            ..G..GC.TG T...C..C.. C.....C..G CAAGTC..A. .C..T...T.
DL20            ..G..TC.TG T...C..C.. C.....T..G CAAGTC..A. .C..T...T.
GA              ..G..GC.TG T...C..C.. C.....T..G CAAGTC..A. .C..T...T.
KU1             ..A..AG.AA C...T..C.. C.....T..G TCTGTC..G. .T..C...C.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus       YCAAAAYCGT TCGAMYWCGT ATAAACCYGG WTAYRTYCGW MRNGRHTTYC
T72             T.....T... ....ATT... .......T.. T..CG.T..A AAG.GA..C.
DL10            T.....T... ....ATA... .......C.. T..TA.C..T CGT.AC..C.
DL20            T.....C... ....ATA... .......T.. A..TA.T..T CGC.AC..T.
GA              T.....T... ....CTA... .......T.. T..CA.C..T CGT.AC..T.
KU1             C.....T... ....CCA... .......T.. T..TA.T..A AAA.AT..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       GWARACCVAC DAACTTYTGG AAYGGCTWTC GCTRTTTCMA TCAGCCMGTY
T72             .T.A...G.. A.....T... ..T....A.. ...A....A. ......C..C
DL10            .A.G...A.. G.....T... ..T....T.. ...A....A. ......A..T
DL20            .A.G...A.. T.....T... ..C....A.. ...G....A. ......A..T
GA              .A.G...A.. G.....C... ..C....A.. ...G....A. ......A..T
KU1             .T.A...C.. G.....T... ..T....A.. ...A....C. ......C..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       GGYRYSTTYA CTCGGAAACT HKMYRAYGGT GGGMGWCAAG WBGCYGATTA
T72             ..CGTC..T. .......... TGACA.C... ...A.A.... TC..T.....
DL10            ..TACC..C. .......... CTCTG.T... ...A.A.... TT..T.....
DL20            ..TACC..C. .......... CTCTG.T... ...A.A.... TT..T.....
GA              ..TACC..C. .......... CTCTG.T... ...A.A.... TC..T.....
KU1             ..CACG..T. .......... ATCTG.T... ...C.T.... AG..C.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       YGGYATHGTR AACCCTAATA AGTTYACBGS YAAYAGYCAG CAYYTKGGRG
T72             C..T..C..G .......... ....C..T.C T..T..T... ..CT.G..A.
DL10            T..C..C..G .......... ....C..G.G C..C..C... ..CT.G..G.
DL20            T..C..T..G .......... ....C..T.C C..C..T... ..CT.G..G.
GA              T..T..C..A .......... ....T..T.C C..T..C... ..CT.G..G.
KU1             T..T..A..G .......... ....C..C.C C..C..T... ..TC.T..A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       ABAACATGGT DATTTAYCCW GGTCCYTTYT CDATHAATAT TGAYMABMGW
T72             .G........ G.....T..A .....T..C. .G..C..... ...TA.TA.A
DL10            .C........ T.....C..T .....C..T. .T..T..... ...TC.CC.A
DL20            .C........ A.....C..T .....C..C. .T..A..... ...TC.GC.T
GA              .T........ A.....C..T .....C..T. .T..A..... ...TC.GC.T
KU1             .G........ T.....T..A .....T..C. .A..A..... ...CA.CA.A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus       GCTWSYGTYG AAGTCCTYAA YAARCTYTCD CARTCNAACC TCAATATTGG
T72             ...AGT..C. .......T.. T..A..T..T ..A..T.... ..........
DL10            ...TCT..C. .......C.. T..A..T..G ..A..A.... ..........
DL20            ...TCT..C. .......C.. T..G..T..A ..G..G.... ..........
```

```
GA          ...TCC..T. .......C.. C..G..C..G ..A..T.... .........
KU1         ...TCT..C. .......C.. T..A..T..T ..A..C.... .........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        560        570        580        590        600
Consensus       GGTTGCDATT GCDGARGCYA ARATGACTGC TTCRYTRCTY KCWMRDCAAT
T72             ......G... ..A..A..T. .A........ ...AC.G..T T.TCGT....
DL10            ......A... ..T..A..T. .G........ ...AT.A..C G.TAAA....
DL20            ......G... ..T..G..T. .G........ ...AT.A..C G.TAAG....
GA              ......A... ..T..A..C. .G........ ...GT.A..C G.TAAA....
KU1             ......T... ..G..A..C. .G........ ...AT.A..C T.ACGT....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        610        620        630        640        650
Consensus       CDATHKCKCT CATCMGAGCC TACACTGCTG CAAAGCGCGG TAADTGGCGR
T72             .G..CT.T.. ....C..... .......... .......... ...G.....G
DL10            .A..AG.T.. ....A..... .......... .......... ...A.....A
DL20            .T..AG.T.. ....A..... .......... .......... ...T.....A
GA              .T..AG.T.. ....A..... .......... .......... ...T.....A
KU1             .T..TG.G.. ....A..... .......... .......... ...G.....A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        660        670        680        690        700
Consensus       GAGGTGYTWT CWCARCTCCT YATYKCCGAA CACCGTTTCA SRRSWCCYKY
T72             ......T.A. .A..A..... T..CG..... .......... CGAGA..CTC
DL10            ......C.T. .T..G..... C..CT..... .......... GAGCT..TGT
DL20            ......T.T. .T..A..... C..CT..... .......... GAGCT..TGC
GA              ......C.T. .T..G..... C..CT..... .......... GAGCT..TGC
KU1             ......C.T. .A..G..... C..TG..... .......... CGAGA..CTC


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        710        720        730        740        750
Consensus       WARGGATCTC GGAGGTCGAT GGCTCGAACT GCAGTACGGY TGGYTWCCYC
T72             A.G....... .......... .......... .........T ...T.A..T.
DL10            T.A....... .......... .......... .........T ...T.A..T.
DL20            T.A....... .......... .......... .........T ...T.A..C.
GA              T.A....... .......... .......... .........C ...C.A..C.
KU1             T.A....... .......... .......... .........T ...C.T..T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        760        770        780        790        800
Consensus       TTATGAGYGA TWTVAARGCT GSHTATGAYY TGCTYACGCA RACYMAWYTR
T72             .......T.. .A.C..G... .GC.....TC ....T..... G..TA.AT.A
DL10            .......C.. .A.G..A... .CA.....TT ....T..... A..CA.AC.G
DL20            .......T.. .T.G..A... .CA.....TT ....C..... A..CA.AC.G
GA              .......T.. .T.G..A... .CA.....TT ....T..... A..CA.AT.G
KU1             .......T.. .A.A..A... .GT.....CT ....T..... A..TC.TT.A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        810        820        830        840        850
Consensus       CCTGCKTTHA TGCCYYTDMG RGTWASYCGY ACCGTTGGCG SHACRCAYAA
T72             .....T..A. ....TC.TA. A..T.CC..C .......... GC..A..C..
DL10            .....T..C. ....CT.GA. G..T.CC..T .......... GA..G..C..
DL20            .....T..C. ....CT.GA. G..T.CC..C .......... GA..G..C..
```

```
GA          .....T..C. ....CT.AA. G..T.CC..C .......... GA..G..C..
KU1         .....G..T. ....TC.TC. A..A.GT..T .......... CT..G..T..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        860        870        880        890        900
Consensus       TTAYAAAGTG CGYAAYGTCG AATCTGCAGG RGATACTTGG TCMTATAGSS
T72             ...T...... ..T..C.... .......... A......... ..A.....CG
DL10            ...T...... ..C..C.... .......... G......... ..C.....CG
DL20            ...T...... ..C..C.... .......... G......... ..C.....CG
GA              ...C...... ..C..C.... .......... G......... ..C.....GC
KU1             ...C...... ..T..T.... .......... G......... ..C.....CG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        910        920        930        940        950
Consensus       AYCGBYTSTC RGTRAATTAC CGAATATGGT ATTTYATYTC YGAYCCVCGV
T72             .T..GC.C.. A..G...... .......... ....C..C.. T..T..C..A
DL10            .T..TT.G.. G..G...... .......... ....C..T.. C..T..A..G
DL20            .T..TT.G.. G..G...... .......... ....C..T.. C..C..G..A
GA              .C..CT.G.. G..G...... .......... ....C..T.. C..C..G..C
KU1             .C..GC.C.. A..A...... .......... ....T..T.. T..T..G..A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        960        970        980        990       1000
Consensus       CTCGCATGGG CHAGTTCYCT HGGKCTYYTW AACCCWYTRG AAATYTAYTG
T72             .......... .A.....T.. A..T..CT.A .....AT.A. ....T..T..
DL10            .......... .T.....C.. T..G..TC.A .....TC.A. ....C..T..
DL20            .......... .T.....C.. T..G..TC.T .....TT.G. ....C..C..
GA              .......... .T.....C.. C..G..TC.T .....TT.A. ....C..T..
KU1             .......... .C.....C.. A..T..CT.A .....TC.A. ....C..T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1010       1020       1030       1040       1050
Consensus       GGARAAGACR CCBTGGTCKT TCGTYGTTGA CTGGTTYYTR CCYGTHGGWA
T72             ...A.....A ..G.....G. ....T..... ......TC.G ..C..A..A.
DL10            ...G.....G ..T.....T. ....C..... ......CT.A ..T..T..T.
DL20            ...A.....G ..C.....G. ....T..... ......CT.A ..T..C..T.
GA              ...G.....G ..C.....G. ....T..... ......CT.A ..T..C..T.
KU1             ...A.....A ..G.....G. ....T..... ......TC.G ..T..C..A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1060       1070       1080       1090       1100
Consensus       ATCTKATMGA AGCYATGAGY AAYCCKCTYG GCYTMGAYAT HATTTCYGGS
T72             ....G..C.. ...C.....T ..C..T..T. ..T.A..C.. C.....T..G
DL10            ....T..A.. ...C.....T ..T..T..T. ..C.C..T.. T.....C..C
DL20            ....T..A.. ...C.....T ..T..T..T. ..C.C..C.. T.....C..C
GA              ....T..A.. ...C.....T ..T..T..T. ..C.C..T.. T.....C..C
KU1             ....T..C.. ...T.....C ..T..G..C. ..C.A..T.. A.....C..G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1110       1120       1130       1140       1150
Consensus       ACDAARACYT GGCAACTYGA ATCWAARHTK AAYGCRWCSM TTMMVGCDBM
T72             ..A..G..C. .......C.. ...T..AT.G ..C..AA.CA ..AAC..ATC
DL10            ..G..G..C. .......C.. ...A..AC.T ..T..GA.GC ..ACA..TTC
DL20            ..G..G..C. .......T.. ...A..AC.T ..C..GA.GC ..CCG..TCC
```

```
GA          ..G..G..T. .......C.. ...A..AC.T ..T..GA.GC ..CCG..TTC
KU1         ..T..A..C. .......C.. ...T..GA.G ..T..GT.CC ..AAA..GGA
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1160       1170       1180       1190       1200
Consensus       DGGHTGGKYY GGRACWGCAA AGYTRWCTGC ATAYGCGAAW RMVKRYGACA
T72             A..C...TCT ..G..A.... ..T.AT.... ...T.....A GCGTAT....
DL10            G..T...TCT ..A..T.... ..T.AT.... ...C.....A GCATAT....
DL20            G..T...TCT ..A..T.... ..T.GA.... ...T.....A GCGTAT....
GA              G..T...TCT ..A..T.... ..T.GA.... ...C.....A GCATAT....
KU1             T..A...GTC ..A..T.... ..C.AT.... ...C.....T AACGGC....

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1210       1220       1230       1240       1250
Consensus       GRTCDACTTT CTAYTCCTTY CCHACBCCKH TGCCKTAYGT GAAATCYCCA
T72             .G..T..... ...T.....T ..A..G..TA ....T..C.. ......C...
DL10            .G..A..... ...C.....T ..A..T..TA ....T..T.. ......C...
DL20            .A..G..... ...T.....C ..C..C..GT ....G..C.. ......T...
GA              .A..G..... ...T.....T ..C..T..TT ....G..C.. ......C...
KU1             .A..T..... ...T.....T ..T..T..TC ....T..C.. ......T...

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1260       1270       1280       1290       1300
Consensus       CTWAGTGGRC TTCACHTRGC VAAYGCRYTM GCMYTAATYA ACCAACGCCT
T72             ..T.....G. .....A.G.. C..T..GC.C ..AT....T. ..........
DL10            ..T.....G. .....T.A.. G..T..AT.A ..CT....C. ..........
DL20            ..T.....G. .....C.A.. G..T..GT.A ..CT....C. ..........
GA              ..T.....G. .....T.A.. G..T..AT.A ..CT....C. ..........
KU1             ..A.....A. .....A.G.. A..C..AC.C ..CC....T. ..........
                      STOP 1                ORF2
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1310       1320       1330       1340       1350
Consensus       GAAAAGGTAA WWAYGGAGTT AGCCAYATGG CAACTTTACG YAGTTTCGTA
T72             .......... AT-C...... .....C.... .......... T.........
DL10            .......... TA.T...... .....T.... .......... C.........
DL20            .......... TT.C...... .....T.... .......... C.........
GA              .......... TT.C...... .....T.... .......... C.........
KU1             .......... AT-C...... .....C.... .......... T.........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1360       1370       1380       1390       1400
Consensus       CTCGTCGATA ATGGCGGTAC GGGGAATGTT ACTGTCGTTC CTGTTAGCAA
T72             .......... .......... .......... .......... ..........
DL10            .......... .......... .......... .......... ..........
DL20            .......... .......... .......... .......... ..........
GA              .......... .......... .......... .......... ..........
KU1             .......... .......... .......... .......... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1410       1420       1430       1440       1450
Consensus       TGCCAACGGC GTCGCTGAGT GGCTTTCTAA TAACTCDCGC AGTCARGCTT
T72             .......... .......... .......... ......G... .....A....
DL10            .......... .......... .......... ......G... .....A....
DL20            .......... .......... .......... ......A... .....A....
GA              .......... .......... .......... ......G... .....G....
KU1             .......... .......... .......... ......T... .....A....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1460       1470       1480       1490       1500
Consensus       ATCGCGTGAC TGCMAGTTAT CGTGCGTCAG GYGCSGAYAA GCGCAARTAY
T72             .......... ...C...... .......... .C..G..C.. ......A..C
DL10            .......... ...C...... .......... .C..G..C.. ......G..C
DL20            .......... ...C...... .......... .C..G..C.. ......A..C
GA              .......... ...C...... .......... .C..G..C.. ......A..T
KU1             .......... ...A...... .......... .T..C..T.. ......A..C

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1510       1520       1530       1540       1550
Consensus       RCYATYAARC TKGARGTRCC GAARATCGTY ACYCARRBHG TWAAYGGHGT
T72             A.C..C..G. .G..G..A.. ...A.....T ..C..GACA. .T..T..T..
DL10            A.C..C..A. .T..A..A.. ...A.....T ..T..AGTT. .T..C..T..
DL20            A.C..T..A. .T..A..A.. ...A.....T ..T..AGTT. .A..T..A..
GA              G.C..T..A. .T..A..A.. ...A.....T ..C..AGTT. .A..T..T..
KU1             A.T..T..G. .G..A..G.. ...G.....C ..T..AAGC. .T..T..C..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       HGARYTGCCT GKTTCSGCRT GGAAGGCTTW YGCCTCYATY GACCTGACCA
T72             T..GC..... .T...C..A. .........A C.....C..T ..........
DL10            C..GC..... .T...C..A. .........A T.....T..T ..........
DL20            T..GC..... .T...C..A. .........A C.....T..C ..........
GA              T..GC..... .G...C..A. .........A T.....T..C ..........
KU1             A..AT..... .T...G..G. .........T C.....C..T ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1610       1620       1630       1640       1650
Consensus       TCCCKATCTT TGCDGCDACY GAYGAYGTGA CYSTTATTTC CAAGTCRCTC
T72             ....T..... ...G..G..C ..C..C.... .TC....... ......A...
DL10            ....G..... ...T..A..T ..C..T.... .TG....... ......A...
DL20            ....T..... ...T..A..C ..T..T.... .CG....... ......A...
GA              ....T..... ...T..A..C ..C..C.... .TG....... ......G...
KU1             ....T..... ...A..T..C ..C..C.... .TC....... ......A...
                                                  ORF3-T72,KU1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1660       1670       1680       1690       1700
Consensus       GCYGGYCTST TCAARRTTGG GRACCCKRTY GCYGAHGCYA TCTCTTCDCA
T72             ..T..T..C. ....GA.... .A....GG.T ..T..T..C. .......A..
DL10            ..T..T..G. ....AG.... .A....TA.C ..T..A..T. .......G..
DL20            ..T..C..G. ....AG.... .G....TA.C ..C..C..T. .......T..
GA              ..C..C..G. ....AG.... .A....TA.C ..T..A..T. .......A..
KU1             ..T..T..C. ....GA.... .A....GG.T ..T..T..C. .......A..
                           STOP 2                            ORF3-DL10,DL20,GA
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1710       1720       1730       1740       1750
Consensus       GAGYGGSTTC TACGCSTAAG GCGTAGTTCT TCAGCATTGG GTCWGAAAGC
T72             ...T..G... .....C.... .......... .......... ...A......
DL10            ...C..C... .....G...- ---------- -------... ...T......
DL20            ...T..C... .....G...- ---------- -------... ...T......
GA              ...T..C... .....G...- ---------- -------... ...T......
KU1             ...T..G... .....C.... .......... .......... ...A......
```

```
                                    ORF4
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1760       1770       1780       1790       1800
Consensus        AWTTAAACAT AAGGAAAACC TATGTTCCGA TTCASAGAGA TCRAAAAGAC
T72              .A........ .......... .......... ....C..... ..A.......
DL10             .---...... .......... .......... ....C..... ..G.......
DL20             .---...... .......... .......... ....C..... ..G.......
GA               .---...... .......... .......... ....G..... ..G.......
KU1              .T........ .......... .......... ....C..... ..G.......

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1810       1820       1830       1840       1850
Consensus        TCTATGTATG GATCGCACTC GCGATTGTGC TGTCCGATTT CACGTCTATC
T72              .......... .......... .......... .......... ..........
DL10             .......... .......... .......... .......... ..........
DL20             .......... .......... .......... .......... ..........
GA               .......... .......... .......... .......... ..........
KU1              .......... .......... .......... .......... ..........

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1860       1870       1880       1890       1900
Consensus        TTCAGTCATT GGATWTGGGG TCTTCTGATC CTCWWTCTCC RGACTTTGAT
T72              .......... ....T..... .......... ...AT..... A.........
DL10             .......... ....A..... .......... ...AT..... A.........
DL20             .......... ....T..... .......... ...AT..... G.........
GA               .......... ....T..... .......... ...TA..... A.........
KU1              .......... ....T..... .......... ...AT..... G.........
                                  STOP 3
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1910       1920       1930       1940       1950
Consensus        GGACTTGCCT ACCTTCGTGA YGARTGTYTA ACTAAACATC CBTCACTHGG
T72              .......... .......... C..A...T.. .......... .T.....T..
DL10             .......... .......... T..G...T.. .......... .T.....A..
DL20             .......... .......... T..A...T.. .......... .C.....C..
GA               .......... .......... T..A...C.. .......... .T.....T..
KU1              .......... .......... T..A...T.. .......... .G.....T..

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1960       1970       1980       1990       2000
Consensus        RRACAGTAAT TCGGAYGCAC KCCGTAAGGA AYTGGCATAT GCCAAACTTA
T72              AG........ .....T.... T......... .T........ ..........
DL10             AA........ .....T.... G......... .T........ ..........
DL20             AG........ .....C.... G......... .T........ ..........
GA               AG........ .....C.... G......... .T........ ..........
KU1              GG........ .....C.... T......... .C........ ..........

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         2010       2020       2030       2040       2050
Consensus        TGGATAGTGA TCAAAGATGC AAAATCCAAA ACAGTAACGG ATACGACYWV
T72              .......... .......... .......... .......... .......CTG
DL10             .......... .......... .......... .......... .......TAC
DL20             .......... .......... .......... .......... .......TAC
GA               .......... .......... .......... .......... .......TAC
```

KU1 ....CTA

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2060       2070       2080       2090       2100
Consensus       WSTCAYATCG ANDSWRGCGT ACTKARYGGH ATAYTMMWRA CYGCYMAGGC
T72             AG...T.... .CTCTG.... ...T.AC..C ...T.ACTA. .C..TA....
DL10            TC...C.... .AAGTA.... ...G.GT..T ...C.CAAG. .T..TC....
DL20            TC...T.... .GAGTG.... ...G.GT..T ...C.CAAG. .T..TC....
GA              TC...T.... .GAGTG.... ...T.GC..T ...C.CAAG. .C..CC....
KU1             AG...T.... .TGCAG.... ...T.AC..A ...T.ACTA. .C..TA....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2110       2120       2130       2140       2150
Consensus       YYYWRTRGCR AASTTRCTTA YRGGYTTTGA ATCTCACTTC CTGAACGATT
T72             TTCAA.A..G ..G..A.... TG..C..... .......... ..........
DL10            CCTTG.G..G ..C..G.... CG..T..... .......... ..........
DL20            CCTTG.G..A ..C..G.... CG..T..... .......... ..........
GA              CCTTG.G..A ..C..A.... CG..T..... .......... ..........
KU1             TTTAA.A..G ..G..A.... TA..T..... .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2160       2170       2180       2190       2200
Consensus       GTTCRTTCTC CAAYGGAGCC TCACAAGGGT TCAAGTTGCR GGATGCAGCG
T72             ....A..... ...T...... .......... .........A ..........
DL10            ....A..... ...T...... .......... .........G ..........
DL20            ....G..... ...C...... .......... .........G ..........
GA              ....A..... ...C...... .......... .........G ..........
KU1             ....A..... ...T...... .......... .........G ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2210       2220       2230       2240       2250
Consensus       CCGTTTAAGA AGATCGCTGG GCAAGCAACC GTTACAGCTC CTGCYTATRA
T72             .......... .......... .......... .......... ....T...G.
DL10            .......... .......... .......... .......... ....T...A.
DL20            .......... .......... .......... .......... ....C...G.
GA              .......... .......... .......... .......... ....T...G.
KU1             .......... .......... .......... .......... ....T...G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2260       2270       2280       2290       2300
Consensus       YHTWGCCGTG SHCGCMGTTA AAACKTGCGS RCCRTGGYWT SSTTATATGC
T72             CC.A...... CA...A.... ....T....G A..A...CT. CG........
DL10            CA.T...... GC...C.... ....G....C G..G...TA. GC........
DL20            CA.T...... GC...C.... ....G....C A..G...TA. GC........
GA              CA.T...... GC...C.... ....G....C A..G...TA. GC........
KU1             TT.A...... CT...A.... ....T....G A..A...CT. CG........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2310       2320       2330       2340       2350
Consensus       AGGAAACGTA TGGCGATGAR ACCARATGGT TTCGTCGGGT RTACGGMAAC
T72             .......... .........G ....G..... .......... A.....A...
DL10            .......... .........A ....G..... .......... A.....C...
DL20            .......... .........A ....A..... .......... G.....C...
```

```
GA          .......... .........G ....A..... .......... A.....C...
KU1         .......... .........G ....G..... .......... G.....C...
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2360       2370       2380       2390       2400
Consensus       GGTTTGTTTT CTGTTCCGAA GAACAATAAA ATAGATCGGG CTGCCTGYAA
T72             .......... .......... .......... .......... .......T..
DL10            .......... .......... .......... .......... .......T..
DL20            .......... .......... .......... .......... .......T..
GA              .......... .......... .......... .......... .......T..
KU1             .......... .......... .......... .......... .......C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2410       2420       2430       2440       2450
Consensus       GGAGCCTGAT ATGAATATGT AYCTWCAGAA RGGKGCKGGM TCTTTTATWA
T72             .......... .......... .T..T..... A..T..T..C ........T.
DL10            .......... .......... .C..T..... G..G..G..A ........A.
DL20            .......... .......... .C..T..... A..G..G..A ........A.
GA              .......... .......... .C..T..... G..G..G..A ........A.
KU1             .......... .......... .T..A..... A..T..T..C ........A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2460       2470       2480       2490       2500
Consensus       GAMRMCGMCT YCGMTCYGTB RRYATWGATC TTAAYGATCA RACRYGCAAY
T72             ..CGC..A.. C..A..T..G AAC..T.... ....C..... A..AC....T
DL10            ..AAA..C.. T..C..C..C GGT..A.... ....C..... A..GT....T
DL20            ..AAA..C.. T..C..C..T GGT..A.... ....T..... A..GC....C
GA              ..AAA..C.. T..C..C..C GGT..A.... ....C..... G..GC....T
KU1             ..CGC..A.. C..A..T..G AAC..T.... ....C..... G..AC....T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2510       2520       2530       2540       2550
Consensus       CAGGARTTRG CCCGACTTGG CAGCATTGAT GGKTCRCTYG CTACTATWGA
T72             .....A..G. .......... .......... ..T..G..T. .......T..
DL10            .....A..A. .......... .......... ..T..A..C. .......T..
DL20            .....A..A. .......... .......... ..G..G..C. .......T..
GA              .....A..A. .......... .......... ..T..G..C. .......T..
KU1             .....G..A. .......... .......... ..T..G..T. .......A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2560       2570       2580       2590       2600
Consensus       TYTWAGTAGY GCTAGYGAYT CYRTCTCTGA YCGYCTYGTM TGGGATYTAC
T72             .T.A.....T .....C..T. .CG....... T..C..C..C ......T...
DL10            .C.T.....C .....T..T. .TG....... C..C..T..C ......T...
DL20            .C.T.....C .....T..C. .CA....... C..T..C..C ......T...
GA              .C.T.....C .....C..T. .CA....... C..T..T..C ......C...
KU1             .T.A.....C .....C..T. .CG....... C..C..C..A ......T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2610       2620       2630       2640       2650
Consensus       TTCCRCCRCA YGTYTATTCA TAYCTCSMYC GYATYCGHWC RTCKTTCACW
T72             ....G..G.. T..T...... ..T...CAC. .T..C..CT. A..T.....T
DL10            ....G..A.. C..T...... ..C...GCT. .C..C..TT. A..G.....T
DL20            ....A..G.. C..C...... ..C...GCT. .T..T..CT. G..T.....T
```

```
GA          ....G..G.. C..T...... ..C...GCT. .T..C..AA. A..G.....T
KU1         ....G..G.. T..T...... ..T...CAC. .C..T..CT. A..T.....A
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2660       2670       2680       2690       2700
Consensus       ATGATYGAYG GDCRBYTRCA YAAGTGGRRY CTRTTTTCTA CBATGGGWAA
T72             .....C..T. .G.AGT.A.. T......AAC ..A....... .G.....T..
DL10            .....C..T. .T.GGT.G.. C......GGC ..G....... .T.....A..
DL20            .....T..C. .G.GCT.A.. C......GGT ..A....... .C.....T..
GA              .....C..T. .G.GTT.A.. T......GGT ..A....... .C.....T..
KU1             .....T..T. .A.GTC.G.. T......AAC ..A....... .G.....T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2710       2720       2730       2740       2750
Consensus       YGGYTTYACR TTCGARCTCG AGTCCATGAT HTTYTGGGCY YTDAGYAASA
T72             C..T..T..A .....G.... .......... A..T.....T C.G..T..G.
DL10            C..C..T..A .....A.... .......... T..C.....T T.G..T..G.
DL20            C..C..T..A .....A.... .......... C..T.....C C.G..T..G.
GA              T..C..C..G .....A.... .......... C..T.....T T.A..C..G.
KU1             C..T..T..G .....A.... .......... A..T.....T C.T..C..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2760       2770       2780       2790       2800
Consensus       SYRTWATGYY GTMCMTSGGT GTTACTGGCT YAYTWGGCRT CTAYGGDGAY
T72             GTG.A...TC ..A.C.C... .......... T.C.T...A. ...T..A..C
DL10            GTG.T...CT ..C.A.G... .......... C.T.A...A. ...C..T..T
DL20            GCG.T...CT ..C.A.G... .......... C.T.A...G. ...C..T..C
GA              GCA.T...CT ..C.A.G... .......... C.T.A...A. ...C..T..T
KU1             CTG.T...TC ..A.C.C... .......... T.C.T...A. ...T..G..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2810       2820       2830       2840       2850
Consensus       GATATAATCG TYCCYRYKRA RTGYSSHCCV MYYCTYYTWM AGGTWYTATC
T72             .......... .C..CACGA. A..CGCA..C CTT..TC.AC ....AT....
DL10            .......... .C..CGTTG. G..TGCC..A ACT..CC.TA ....AC....
DL20            .......... .T..TGTTG. A..TGCC..G ACC..CC.TA ....AT....
GA              .......... .C..CGTTG. G..TCGT..A ACT..CC.TA ....AC....
KU1             .......... .T..CACGA. A..CGCA..C CTT..CT.AC ....TT....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2860       2870       2880       2890       2900
Consensus       SGCYGTRAAY TTTCTTCCKA AYVAGRAGAA RACRTTYACD ACSGGKTAYT
T72             G..T..G..C ........T. .CC..A.... G..G..T..T ..C..T..T.
DL10            C..T..G..C ........T. .TA..A.... G..G..C..T ..G..T..C.
DL20            C..T..G..C ........G. .TC..A.... G..G..T..G ..G..T..C.
GA              C..T..A..C ........T. .TG..G.... A..A..T..A ..G..T..C.
KU1             G..C..A..T ........T. .CC..A.... A..A..C..G ..C..G..T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2910       2920       2930       2940       2950
Consensus       TYCGTGARAG TTGYGGKGCH CAYTTCTTYA AAGRYGCYKM MRTRAAACCT
T72             .C.....G.. ...T..G..A ..T.....T. ...GT..TTC AG.A......
DL10            .T.....A.. ...T..T..T ..C.....T. ...GT..CGA CA.G......
DL20            .T.....G.. ...C..T..T ..T.....T. ...GC..CGA CA.G......
```

```
GA          .T.....A.. ...T..T..C ..C.....C. ...AT..CGA CA.G......
KU1         .C.....G.. ...T..G..A ..T.....C. ...GT..TTC AG.A......
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2960       2970       2980       2990       3000
Consensus       TTTTACTGCA AGCGGCCRAT GGAAACCCTT CCSGAYRTHM TRTTRCTMTG
T72             .......... .......G.. .......... ..C..TA.TA .G..A..C..
DL10            .......... .......A.. .......... ..C..TG.CA .G..G..A..
DL20            .......... .......A.. .......... ..C..TG.CA .G..G..C..
GA              .......... .......A.. .......... ..C..TG.CA .G..G..A..
KU1             .......... .......G.. .......... ..G..CG.AC .A..A..A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3010       3020       3030       3040       3050
Consensus       YAAYAGRATH MGDGGTTGGS RKACYRTYGG KGGAATRTCA GAYCCKCGRC
T72             T..C..A..T C.T......G GT..TA.C.. G.....A... ..C..T..A.
DL10            C..C..G..A A.G......C AG..CG.T.. T.....G... ..T..G..G.
DL20            C..C..G..A A.G......C AG..CG.T.. T.....G... ..T..G..G.
GA              C..C..G..A A.A......C AG..CG.T.. T.....G... ..T..G..A.
KU1             T..T..G..C C.G......G GG..TA.C.. G.....A... ..C..T..G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3060       3070       3080       3090       3100
Consensus       TCTTTCCDAT WTGGAARGAG TTYGCWGATA TGATHCCGCC KAAGTTTAAG
T72             .......A.. A.....G... ..T..A.... ....A..... T.........
DL10            .......T.. T.....G... ..T..T.... ....T..... G.........
DL20            .......G.. T.....G... ..T..T.... ....T..... G.........
GA              .......T.. T.....G... ..C..T.... ....C..... G.........
KU1             .......G.. A.....A... ..T..A.... ....A..... T.........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3110       3120       3130       3140       3150
Consensus       GGTGGATGYA ACCTKGATCG CGAYACTTAC CTYGTYTCAC CKGATAAACC
T72             ........T. ....T..... ...C...... ..T..T.... .G........
DL10            ........C. ....G..... ...T...... ..C..T.... .G........
DL20            ........T. ....G..... ...T...... ..C..T.... .T........
GA              ........C. ....G..... ...T...... ..C..C.... .T........
KU1             ........T. ....G..... ...C...... ..T..C.... .G........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3160       3170       3180       3190       3200
Consensus       TGGTRWHWCA YTRGTTCGWR WYGCKAMGRW RCGHWSTGGT TTTAACYAYD
T72             ....AAAA.. T.A.....TG TT..G.A.AA A..AAG.... ......C.TA
DL10            ....GTCA.. T.G.....AG TC..T.A.GT A..TTC.... ......C.TT
DL20            ....GTCA.. C.G.....TA TT..T.A.GT A..CTC.... ......C.CG
GA              ....GTTT.. T.G.....TA TT..T.A.GT A..CTC.... ......C.TG
KU1             ....GTAA.. T.A.....TG AT..G.C.GT G..TAG.... ......T.CA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3210       3220       3230       3240       3250
Consensus       MRTTYCSKWR BSRCYAWGAA AAYGGYCGYT AYRTYCAYTG GYTRCATATG
T72             AG..T.GTAG TGA.T.T... ..T..T..C. .TA.T..T.. .T.A......
DL10            CG..C.CGTA CGG.C.T... ..T..T..C. .CG.C..C.. .T.G......
DL20            CG..C.CGTA CGG.T.T... ..C..C..T. .CG.C..T.. .C.A......
```

```
GA          CG..C.CGTA CGG.C.T... ..T..T..C. .CG.C..C.. .T.G......
KU1         AA..C.GTAG GCG.C.A... ..T..C..C. .TA.C..T.. .T.A......
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3260       3270       3280       3290       3300
Consensus       GGATCGGGHG AAGTDYYYGA RACYATWTCY TCCGCCAGGT WTAGGTGCAA
T72             ........C. ....GCTC.. A..T..A..T .......... T.........
DL10            ........A. ....TCTC.. G..C..A..C .......... T.........
DL20            ........A. ....TCTC.. G..C..A..C .......... T.........
GA              ........A. ....TCTT.. G..C..T..C .......... A.........
KU1             ........T. ....ATCC.. A..T..A..T .......... T.........
                                                                STOP 4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3310       3320       3330       3340       3350
Consensus       ACCTAAYTCG GAATGGAGRA YACAGATCCC KCTATTTCCT CAGGAADTAG
T72             ......T... ........G. C......... G......... ......A...
DL10            ......C... ........A. C......... T......... ......T...
DL20            ......C... ........A. C......... T......... ......T...
GA              ......C... ........A. C......... T......... ......T...
KU1             ......C... ........A. T......... T......... ......G...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3360       3370       3380       3390       3400
Consensus       AGGCCTGCGT TCTCTCCTGA TAGTATCAGG ACCTCCCCGK AYGGGGTGGG
T72             .......... .......... .......... .........G .T.......
DL10            .......... .......... .......... .........G .T.......
DL20            .......... .......... .......... .........G .C.......
GA              .......... .......... .......... .........T .T.......
KU1             .......... .......... .......... .........G .T.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3410       3420       3430       3440       3450
Consensus       TGKGACCGWA AGGYYMCTAT GGRGGTGAWH YCYYCCACCG CCBWWAWTGG
T72             ..G.....A. ...TTC.... ..A.....TA T.CT.--... ..GTA.--..
DL10            ..T.....A. ...CCA.... ..G.....AT C.CC.-.... ..CAA.A-..
DL20            ..T.....A. ...CCA.... ..A.....AC C.TC.-.... ..CAA.A-..
GA              ..T.....A. ...CCA.... ..A.....AC C.TC...... ..CAA.A-..
KU1             ..G.....T. ...TTC.... ..A.....TA C.CT.--... ..TTT.T...


                ....|....| ....|....| ....|....| ....|....|
                        3460       3470       3480       3490
Consensus       CGGTTYWMGG TGACTAGTTT GCTTGGCTAG TCACCACCCA
T72             ..-..TAA.. .......... .......... ..........
DL10            .....TTA.. .......... .......... ..........
DL20            .....CTC.. .......... .......... ..........
GA              .....CTC.. .......... .......... ..........
KU1             .....TAA.. .......... .......... ..........
```

Appendix A3a
Group III

Alignment: Allolevivirus Group III.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       GGGKKHCCCC CCKTAGGGGG KYWCYYYAYR YAGYAGTAYT TCAMYVAKTA
TW18            ...GGA.... ..T....... TCA.CTC.CA C..C....C. ...CTA.G..
HL4-9           ...GGA.... ..T....... TCA.CTC.CA C..C....T. ...CTG.G..
BR12            ...GGA.... ..T....... TCA.CTC.CA C..C....T. ...CTG.G..
BZ1             ...GGA.... ..T....... TCT.CTC.CA C..C....T. ...CTG.G..
VK              ...GGA.... ..T....... TCA.CTC.CA C..C....T. ...CTG.G..
QB              ---------- ---------- ---------- ---------- -..CTG.G..
M11             ...TTC.... ..G....... GTA.TCT.TG T..T..---- -..ACC.T..
MX1             -..GTT.... ..G....... GTA.TCT.TG T..T..---- -..ACC.T..
                         ORF1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       YDRGAGGMMA YATGCCWMRW YTACCDMGKG SWCTKCGWTT CGGASCSRAY
TW18            TGA....AC. T.....TAAA T....AC.T. GT..G..T.. ....G.CG.T
HL4-9           TGA....AC. T.....TAAA T....AC.T. GT..G..T.. ....G.CG.T
BR12            TAA....AC. T.....TAAA T....AC.T. GT..G..T.. ....G.CG.T
BZ1             TAA....AC. T.....TAAA T....GC.T. GT..G..T.. ....G.CG.T
VK              TGA....AC. T.....TAAA T....AC.T. GT..G..T.. ....G.CG.T
QB              TAA....AC. T.....TAAA T....GC.T. GT..G..T.. ....G.CG.T
M11             CTG....CA. C.....ACGT C....TA.G. GA..T..A.. ....G.GA.C
MX1             CTG....CA. C.....ACGT T....TA.G. CA..T..T.. ....C.GA.T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       AWKGARRTYY TWARYGAYTT YCARGARCTC TGGTWTCCAG ABYYCHKYRT
TW18            .AT..AA.TC .T.AT..T.. T..G..G... ....T..... .CCT.TTTA.
HL4-9           .AT..AA.TC .T.AT..T.. T..A..G... ....T..... .TCT.TTTA.
BR12            .AT..AA.TC .T.AT..T.. T..G..G... ....T..... .TCT.TTTA.
BZ1             .AT..AA.TC .T.AT..T.. T..G..G... ....T..... .TCT.TTTA.
VK              .AT..AA.TC .T.AT..T.. T..G..G... ....T..... .TCT.TTTA.
QB              .AT..AA.TC .T.AT..T.. T..G..G... ....T..... .CCT.TTTA.
M11             .TG..GG.CT .A.AT..C.. C..G..A... ....A..... .GTC.CGCG.
MX1             .TG..GG.CT .A.GC..C.. C..G..A... ....A..... .GTC.ATTA.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       CGAWTCYKMY RHSAHDYWYC CKTKGTAYAC MYTSARRGGT MRYRTSKKKR
TW18            ...A..TTCC GAC.CGCAT. .G.G...C.. AC.G.AA... CGTG.GTTGA
HL4-9           ...A..TTCC GAC.CGCAT. .G.G...C.. AC.G.AA... CGTG.GTTGA
BR12            ...A..TTCC GAC.CGCAT. .G.G...C.. AC.G.AA... CGTG.GTTGA
BZ1             ...A..TTCC GAC.CGCAC. .G.G...C.. AC.G.AA... CGTG.GTTGA
VK              ...A..TTCC GAC.CGCAT. .G.G...C.. AC.G.AG... CGTG.GTTGA
QB              ...A..TTCC GAC.CGCAT. .G.G...C.. AC.G.AG... CGTG.GTTGA
M11             ...T..CGAT ACG.TTTTC. .T.T...T.. CT.C.AA... AACA.GGGTG
MX1             ...T..CGAT GTG.AATAC. .T.T...T.. CT.C.GA... AGCA.CGGTG
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
Consensus       RHKCYYWYHT NGAYDMYYRY BKMMCNAAYR WWRKYGKYCG YSARRTHMGD
TW18            ACG.TCATC. T..TGATCGC CTAC.T..TG TAGGT.GT.. CC.GG.CA.G
HL4-9           ATG.CCATC. T..TGATCGT CTAC.T..TG TAGGC.GT.. TC.GA.CA.G
BR12            ACG.TCACA. G..TGACCGC TTAC.T..TG TAAGT.GT.. CC.AG.AC.G
BZ1             ACG.TCATT. A..TGATCGT CTAC.T..TG TAGGC.GT.. TC.AA.AA.G
VK              ACG.TCATA. T..TGACCGT TTAC.C..TG TAAGC.GT.. CC.AG.AA.G
QB              ACG.CCACC. T..TGATCGT CTAC.T..TG TAGGC.GT.. CC.GG.AA.G
M11             GCT.TTTTT. C..CACTTAC GGCA.A..CA ATATT.TC.. TC.GG.TC.A
MX1             GAT.TTTCT. T..TTCTTAC GGCA.G..TA ATATC.TC.. CG.GA.TC.T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus       MGNACDCCDC AYYGYGYYAC HGTHCCKATW GCCWSTWCWG GCYTWMGKCC
TW18            C.T..T..A. .TC.T.TT.. T..T..G..T ...TC.T.A. ..C.TC.T..
HL4-9           C.T..T..A. .TC.T.CT.. T..T..G..T ...TC.T.A. ..C.TC.T..
BR12            C.T..T..A. .TC.C.CT.. T..T..G..T ...TC.A.A. ..C.TC.T..
BZ1             C.T..T..G. .CC.T.CT.. C..C..G..T ...TC.T.A. ..C.TC.T..
VK              C.T..A..A. .TC.C.TT.. T..C..G..T ...TC.T.A. ..C.TC.T..
QB              C.C..T..A. .TC.C.TC.. C..T..G..T ...TC.T.A. ..C.TC.T..
M11             A.G..G..A. .CC.T.CT.. A..T..T..A ...TC.T.T. ..T.AA.G..
MX1             A.A..G..T. .CT.T.CT.. A..A..T..A ...AG.T.T. ..T.AA.G..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       VKKWACMWCY GTWYRGTATG AYCCHDCMRS MYTRYYBTTC AGGATTVYTG
TW18            GGTA..AA.C ..TCA..... .T..CG.AGC AC.ATCG... -----..AT..
HL4-9           GGTA..AA.C ..TCA..... .T..CA.AGC AC.ATCG... -----..AT..
BR12            AGTA..AA.C ..TCA..... .T..CG.AGC AC.ATCG... -----..AT..
BZ1             GGTA..AA.C ..TCA..... .T..TA.AGC AC.GTCG... -----..AT..
VK              AGTA..AA.C ..TCA..... .T..CG.AGC AC.ATCG... -----..GT..
QB              GGTA..AA.C ..TCA..... .T..CG.AGC AC.ATCG... -----..AT..
M11             CTGT..CT.T ..ATG..... .C..AT.CAG CT.GCTT... ......CC..
MX1             CTGT..CT.C ..TTG..... .C..AA.CAG CT.GCTC... ......CC..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       ARCRYBMGDG YKGWBTGGGA YWWYGGYRWK GGYGAYRSTG SRRAYMTYGT
TW18            .A.GCTC.T. TT.AT..... TTTC..TAAT ..C..TAG.. CGA.CC.T..
HL4-9           .A.GCTC.T. TT.AT..... TTTC..TAAT ..C..TAG.. CGA.CC.T..
BR12            .A.GCTC.T. TT.AT..... TTTC..TAAT ..C..CAG.. CGG.CC.T..
BZ1             .A.GCCC.T. TT.AT..... TTTC..TGAT ..T..TAG.. CGG.CC.T..
VK              .A.GCTC.T. TT.AT..... TTTC..TAAT ..C..CAG.. CGG.CC.T..
QB              .A.GCTC.T. TT.AC..... TTTC..TAAT ..C..TAG.. CGA.CC.T..
M11             .A-ATGA.A. CG.TT..... CAAT..CATG ..T..TAC.. GAG.TA.T..
MX1             .G-ATGA.G. CT.AG..... CAAT..CATG ..T..TGC.. GAG.TA.C..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       CWWTAAWGAC TTYBTBTTYM RYACBYYBGC RCCTAARGAK TTYGATTTYT
TW18            .AT...T... ..TT.G..TC GC..TTCC.. A.....G..G ..C.....C.
HL4-9           .AT...T... ..TT.G..TC GT..TTTC.. G.....G..G ..C.....C.
BR12            .AT...A... ..TG.T..CC GT..CTTT.. A.....G..T ..C.....C.
BZ1             .AT...T... ..TT.G..CC GC..TTTC.. A.....G..G ..C.....C.
VK              .AT...A... ..CG.G..TC GC..CTTT.. A.....G..T ..C.....C.
QB              .AT...T... ..TC.G..TC GC..CTTT.. A.....G..G ..T.....T.
M11             .TA...T... ..TC.C..CA AT..GCCG.. A.....G..G ..C.....C.
MX1             .TA...A... ..TC.C..CA GC..CCCG.. A.....A..G ..C.....C.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       CKAAYTCYTT RGYKCCWCGY TAYASYMAKG CCTTCTCYGC BTTTAAYGCY
TW18            .G..T..T.. A.TT..T..T ..T.CTC.G. .......C.. T.....T..C
HL4-9           .G..T..C.. A.TT..T..T ..T.CTC.G. .......C.. T.....T..C
BR12            .G..T..C.. A.CT..T..T ..T.CCC.G. .......C.. C.....T..C
BZ1             .G..C..C.. G.CT..T..C ..T.CCC.G. .......C.. T.....T..C
VK              .G..C..C.. A.TT..T..T ..T.CCC.G. .......C.. T.....T..C
QB              .G..C..C.. A.TT..T..T ..T.CTC.G. .......C.. G.....T..C
M11             .T..T..C.. A.CG..A..C ..C.GTA.T. .......T.. T.....C..T
MX1             .G..T..C.. A.CG..A..C ..C.GTA.T. .......C.. T.....T..C

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus       AARTATGGCR YTATSATYGG HGAAGGGCDC GARACWMTWA ARTATYTCGS
TW18            ..G......A C...G..C.. C.......T. ..G..TA.A. .A...C...G
HL4-9           ..G......A C...G..C.. C.......T. ..G..TA.A. .A...C...G
BR12            ..A......G T...G..C.. T.......T. ..G..TA.A. .A...C...G
BZ1             ..G......A C...G..C.. C.......T. ..G..TA.A. .A...C...G
VK              ..A......G T...G..C.. T.......T. ..G..TA.A. .A...C...G
QB              ..G......A C...G..C.. C.......T. ..G..TA.A. .A...C...G
M11             ..A......G T...C..T.. A.......G. ..G..AC.T. .G...T...C
MX1             ..A......G T...C..C.. T.......A. ..A..AC.T. .G...T...C

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       560        570        580        590        600
Consensus       GCTDTTACTK CGCAGRCTRC RTRARGSWKW CCGYGCYGTT MRRCRYGGMG
TW18            ...G.....G .....A..G. G.G.G.GTTA ...C..T... AAA.GC..C.
HL4-9           ...T.....G .....A..G. G.G.G.GTTA ...C..T... AAG.GT..C.
BR12            ...T.....G .....A..G. G.G.G.GTTA ...C..T... AAG.GC..C.
BZ1             ...T.....G .....A..G. G.G.G.GTTT ...T..T... AAG.AC..C.
VK              ...T.....G .....A..G. G.G.G.GTTA ...T..T... AAG.GC..C.
QB              ...T.....G .....A..G. G.G.G.GTTA ...C..T... AAG.GT..C.
M11             ...A.....T .....G..A. A.A.A.CAGT ...T..C... CGG.AC..A.
MX1             ...G.....T .....G..A. A.A.A.CAGT ...T..C... CGG.AC..A.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        610        620        630        640        650
Consensus       ATTTACGHGS TCTYCGBARR RTYMTYSAST CYTAYMAYAA DGGTMRDTGG
TW18            .......T.C ...T..C.GG G.TA.CC.G. .T..CC.T.. T...AAG...
HL4-9           .......T.C ...T..T.GG G.TA.CC.G. .C..CC.T.. G...AAG...
BR12            .......T.C ...T..C.GG G.TA.CC.G. .T..CC.T.. A...AAA...
BZ1             .......C.C ...T..T.GG G.TA.CC.G. .T..CC.C.. G...AAA...
VK              .......T.C ...T..T.GG G.TA.CC.G. .T..CC.T.. A...AAA...
QB              .......T.C ...T..T.GG G.TA.CC.G. .C..CC.T.. T...AAG...
M11             .......A.G ...C..G.GG A.CC.TG.C. .T..CC.T.. G...CAT...
MX1             .......T.G ...C..G.AA A.CC.CG.C. .T..TA.T.. G...CGT...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        660        670        680        690        700
Consensus       AARCCDRCTA CTGCTGGTAA TCTCTGGCTY GARTTYMGKT AYGGYYTHRY
TW18            ..G..GA... .......... .........T ..A..TC.T. .T..CC.TAT
HL4-9           ..A..AA... .......... .........T ..A..CC.T. .T..TC.TAT
BR12            ..A..GA... .......... .........T ..A..TC.T. .T..CC.TAT
BZ1             ..A..GA... .......... .........T ..A..TC.T. .T..CC.TAT
VK              ..A..GA... .......... .........T ..A..TC.T. .T..CC.TAT
QB              ..A..GG... .......... .........T ..A..TC.T. .T..CC.TAT
M11             ..G..TG... .......... .........C ..A..TC.T. .T..CC.CGT
MX1             ..G..TG... .......... .........C ..G..TA.G. .C..TT.AAC


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        710        720        730        740        750
Consensus       BCCBCTCTTY YAYGACATYA RARRYGTYAT GNWHGAYTGG MMVMRBMKYM
TW18            G..T.....C T.T.....C. A.GAC..C.. .TTA..T... CAGAACCGTC
HL4-9           G..T.....C T.T.....C. A.GAC..C.. .TTA..C... CAGAAGCGCC
BR12            G..T.....C T.C.....C. G.GAT..T.. .TTA..T... CAGAGCCGCC
BZ1             G..C.....C T.T.....C. G.GAT..C.. .TTA..T... CAGAATCGTC
VK              G..T.....C T.T.....C. G.GAT..C.. .CTA..T... CAAAACCGCC
QB              G..T.....T T.T.....C. G.GAT..C.. .TTA..C... CAGAACCGTC
M11             C..G.....C C.T.....C. A.GAT..C.. .AAC..C... ACGCGCATTA
MX1             T..G.....C C.T.....T. A.AGT..T.. .GAT..T... AACCGTATTA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        760        770        780        790        800
Consensus       AYGAYARRAT YCARMRMYWY CKYCGDTTYT CDGTBGGTCA YGGYGAGGAY
TW18            .T..T.AA.. T..ACGCCTC .TT..G..T. .T..C..... C..T.....C
HL4-9           .T..T.GG.. T..ACGCCTC .TT..G..T. .G..T..... C..T.....T
BR12            .T..T.AG.. T..ACGCCTC .TT..A..T. .A..T..... T..T.....C
BZ1             .C..T.AG.. T..ACGCCTC .TT..G..T. .T..C..... T..C.....T
VK              .T..T.AG.. T..ACGCCTC .TT..G..T. .T..T..... C..C.....C
QB              .T..T.AG.. T..ACGCCTC .TT..G..T. .T..T..... C..C.....T
M11             .C..T.AG.. C..GAAATAT .GC..A..C. .G..G..... T..T.....T
MX1             .C..C.AG.. C..GAAACTT .GC..T..C. .T..T..... C..T.....T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       810        820        830        840        850
Consensus       TTYRHGYTRT CDARTTHGAY RRYYTRTAYC CHGSYBTWDC YYAYTTYMRR
TW18            .--GC.T.G. .G.A..C..C AACT.A..T. .A.CCG.TG. TT.C..TAAA
HL4-9           .--AT.T.G. .A.A..C..C AGCT.A..T. .A.CCG.TT. TT.C..TAAA
BR12            .--AC.T.G. .A.G..T..C AACT.G..T. .C.CCT.AG. CT.C..TAAA
BZ1             .--AT.T.G. .G.A..T..C AACT.A..T. .C.CCG.AG. TT.C..TAAA
VK              .--AT.T.G. .A.A..C..C AACT.A..C. .C.CTT.AG. CT.C..TAAA
QB              .--AC.T.G. .G.A..C..C AATC.G..C. .T.CCG.TG. TT.C..TAAA
M11             ..TAA.C.G. .T.--.C..C GGCT.A..T. .T.GCC.TA. TC.T..CCGA
MX1             ..CAA.C.A. .A.--.A..T GGTT.G..T. .T.GCC.AA. CC.C..CAGG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       860        870        880        890        900
Consensus       YTRWVHGGBG AGATTACHST CSARCGCCGT CATCGDYRBG GYATADYYTA
TW18            C.GAAA..T. .......AC. .G.A...... .....TCAT. .T...TCT..
HL4-9           C.AAGA..G. .......AC. .G.A...... .....TCAT. .T...TCT..
BR12            C.GAAA..T. .......TC. .G.A...... .....TCAT. .C...TCT..
BZ1             C.GAAA..T. .......TC. .G.A...... .....TCAT. .T...TCT..
VK              C.GAAA..T. .......AC. .G.A...... .....TCAC. .T...TCC..
QB              C.GAAA..G. .......AC. .G.A...... .....TCAT. .C...TCT..
M11             T.ATCT..T. .......CG. .C.G...... .....ATGG. .T...GTC..
MX1             T.ATCC..C. .......CG. .C.G...... .....GTGG. .T...ACC..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       910        920        930        940        950
Consensus       CGCKAAYCGC GRRGGHTAYG CYRYWTTYGA YAACGGTTCC MTTCGGCCYG
TW18            ...T..T... .AA..A..T. .TGTT..C.. T......... C.......T.
HL4-9           ...T..T... .AG..A..T. .TGTT..C.. T......... C.......T.
BR12            ...G..T... .AA..A..T. .TGTT..C.. T......... C.......T.
BZ1             ...T..T... .AA..A..T. .TGTT..C.. T......... C.......T.
VK              ...G..T... .GG..A..T. .TGTT..C.. C......... C.......T.
QB              ...T..C... .AA..A..T. .TGTT..C.. C......... C.......T.
M11             ...T..T... .AA..C..C. .CACA..T.. C......... A.......C.
MX1             ...T..T... .AA..T..C. .CACA..C.. C......... A.......C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       960        970        980        990       1000
Consensus       TGTCCGAYTG GAAGGARCTY GCYRHYGCDT TYATCAAYCC KSRHGAAGTT
TW18            .......T.. ......G..T ..CACC..A. .C.....T.. GCAT......
HL4-9           .......T.. ......A..T ..CACT..A. .C.....T.. TCAT......
BR12            .......T.. ......A..T ..CGTC..G. .C.....T.. TCAC......
BZ1             .......T.. ......G..T ..CACC..G. .C.....T.. TCAA......
VK              .......T.. ......G..T ..CGTC..G. .C.....T.. TCAC......
QB              .......T.. ......G..T ..CACT..A. .C.....T.. GCAT......
M11             .......C.. ......A..C ..TAAC..T. .T.....T.. TGGC......
MX1             .......C.. ......A..C ..TAAC..G. .T.....C.. TGGC......
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1010       1020       1030       1040       1050
Consensus       GCNTGGGARY TAACWCCHTA CAGCTTYRTY GYBGATTGGT TYWTSAAYGT
TW18            ..T.....GC ....T..C.. ......CG.T .TT....... .CT.G..T..
HL4-9           ..A.....GT ....T..C.. ......CG.T .TT....... .TT.G..T..
BR12            ..T.....GT ....T..T.. ......CG.T .CC....... .TT.G..C..
BZ1             ..C.....GT ....T..C.. ......TG.T .TT....... .TT.G..C..
VK              ..T.....GT ....T..T.. ......TG.T .TT....... .CT.G..C..
QB              ..T.....GT ....T..C.. ......CG.T .TT....... .CT.G..T..
M11             ..G.....AT ....T..A.. ......CG.C .TT....... .TA.C..C..
MX1             ..A.....AT ....A..A.. ......CA.C .TG....... .TA.C..C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1060       1070       1080       1090       1100
Consensus       YGGYGAYATH MTYGMKCARC ARRRKCARYK RTATCAKAAY ATYGABATYG
TW18            C..T..C..A C.T.CT..A. .GGGT..GCT A.....T..T ..C..T..T.
HL4-9           C..T..C..A C.T.CT..G. .GGGT..GCT A.....T..T ..C..T..T.
BR12            T..C..C..A C.T.CT..G. .AGGT..GCT A.....T..T ..C..G..C.
BZ1             C..T..C..A C.C.CT..A. .AGGT..ACT A.....T..T ..C..T..T.
VK              C..C..C..A C.T.CT..G. .AGGT..ACT A.....T..T ..C..T..T.
QB              T..T..C..A C.T.CT..A. .AGGT..GCT A.....T..T ..C..T..T.
M11             T..C..T..C A.C.AG..G. .GAAG..ATT G.....G..C ..C..C..T.
MX1             T..C..T..T A.T.AG..G. .GAAG..ATG G.....G..T ..T..C..T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1110       1120       1130       1140       1150
Consensus       THGAYGGHTW YSASMGWCGY GAYATHCGNH TSMRHTCNKT YWCYMTHAAA
TW18            .A..C..T.T TG.CA.A..T ..C..C..GC .CAAA..CT. CA.TA.A...
HL4-9           .A..C..T.T TG.CA.A..T ..C..T..AC .CAAA..TT. CA.TA.A...
BR12            .A..T..T.T CG.CA.A..T ..C..A..GT .CAAA..AT. CA.TA.T...
BZ1             .A..C..C.T CG.CA.A..T ..C..A..GT .CAAA..GT. TA.CA.A...
VK              .A..T..T.T TG.CA.A..T ..C..A..GT .CAAA..GT. TA.CA.T...
QB              .A..C..C.T TG.CA.A..T ..C..C..GC .CAAA..TT. CA.CA.A...
M11             .T..C..A.A CC.GC.A..T ..T..A..TA .GCGC..CG. TA.TC.C...
MX1             .C..C..T.A CC.GC.T..C ..T..A..CA .GCGT..CG. CT.TC.T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1160       1170       1180       1190       1200
Consensus       GGWGWRCGDA ATGGVMDRCC TGTWMRCGTW WCTGCBRRYY WRTCKRACYT
TW18            ..T.AA..A. ....GCGG.. ...TAA...T T....TGACC TG..TG--C.
HL4-9           ..T.AA..A. ....GCGA.. ...TAA...T T....CGACC TG..TG--C.
BR12            ..T.AA..A. ....ACAG.. ...TAA...A T....GGACT TA..TA--C.
BZ1             ..T.AG..A. ....ACAG.. ...TAA...T T....GGACC TG..TG--C.
VK              ..T.AA..A. ....ACAG.. ...TAA...A T....GGACC TA..TG--C.
QB              ..T.AA..A. ....GCGG.. ...TAA...T T....TAGCC TG..TG--C.
M11             ..A.TA..T. ....CATA.. ...ACG...T A...--GATC AG..GA..T.
MX1             ..A.TA..G. ....CATA.. ...ACG...T A...--GATC AG..GA..T.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1210       1220       1230       1240       1250
Consensus       RTYGATWYMT TTTAYARYCG MYYYCATRCK ASYMRBMTYC CGYWMGCYAC
TW18            G.C...TTA. ....C.GC.. ACTC...A.G .GCAGTC.T. ..TTC..T..
HL4-9           G.C...TCA. ....C.GC.. ACTC...A.G .CCAAGC.T. ..TTC..T..
BR12            A.C...ACA. ....T.GT.. ACTT...G.T .GCAATA.T. ..TTC..T..
BZ1             G.T...ACA. ....T.GT.. ACTC...A.G .GCAGTA.T. ..TTC..T..
VK              A.C...ACA. ....T.GT.. ACTT...G.T .GCAACA.T. ..TTC..T..
QB              G.C...TTA. ....C.GC.. ACTC...A.G .GCAATC.T. ..TTC..T..
M11             G.T...TCC. ....C.AT.. CTCT...A.T .CTCGTA.C. ..CAA..C..
MX1             G.T...TCC. ....C.AT.. CTCT...A.T .CTCGTA.C. ..CAA..C..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1260       1270       1280       1290       1300
Consensus       ACTHGMWMTY GATACDWCYT TYWSKWSBWW TAARCACGTB HTNGAYAGTR
TW18            ...T.ATC.C .....TA.C. .TAGTTCGTA ...A.....C C.T..T...A
HL4-9           ...C.ATC.C .....TA.C. .TAGTTCGTA ...A.....T C.T..T...A
BR12            ...A.ATC.C .....TA.C. .TAGTTCGTT ...A.....C C.A..T...A
BZ1             ...A.ATC.C .....TA.C. .TAGTTCGTT ...A.....T C.C..T...A
VK              ...T.ATC.C .....TA.C. .TAGTTCGTT ...A.....T C.A..T...G
QB              ...A.ATC.T .....TA.C. .TAGTTCGTT ...A.....T C.T..T...A
M11             ...A.AAC.C .....AT.T. .CTCGAGCAT ...G.....G T.A..T...A
MX1             ...A.CAA.C .....GT.C. .CTCGAGTAT ...A.....G A.G..C...A
                                                 STOP 1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1310       1320       1330       1340       1350
Consensus       TYKYTYTWWT AACYCARCGC RTWAARCGYT GAMMWCWTTG GGTCAATTHG
TW18            .CTT.T.AT. ...C..A... A.T..G..T. ..AA-.T... ........T.
HL4-9           .CTT.T.AT. ...C..A... A.T..G..T. ..AA-.T... ........T.
BR12            .CTT.T.AT. ...C..A... A.T..G..T. ..AA-.T... ........T.
BZ1             .CTT.T.AT. ...C..A... A.T..G..C. ..AAT.T... ........T.
VK              .CGC.T.AT. ...C..A... A.T..G..T. ..AA-.T... ........T.
QB              .CTT.T.AT. ...C..A... G.A..G..T. ..AA-.T... ........T.
M11             .CTC.C.TA. ...T..G..N A.A..A..C. ..CCA.A... ........A.
MX1             .TTC.C.TA. ...T..G... A.A..A..C. ..CCA.A... ........C.
                  ORF2/3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1360       1370       1380       1390       1400
Consensus       ATCATGGCWA AATTASARRC TRTYACTTTA RGTRRYATYG GGAARRAHGG
TW18            ........A. .....G.GA. .G.T...... A..AAC..C. ....AG.T..
HL4-9           ........A. .....G.GA. .G.T...... A..AAC..C. ....AG.T..
BR12            ........A. .....G.GA. .G.T...... A..AAC..C. ....AG.T..
BZ1             ........A. .....G.GA. .G.T...... A..AAC..T. ....AG.T..
VK              ........A. .....G.GA. .G.T...... A..AAC..C. ....AG.T..
QB              ........A. .....G.GA. .G.T...... G..AAC..C. ....AG.T..
M11             ........T. .....C.AG. .A.C...... A..GGT..T. ....GA.A..
MX1             ........T. .....C.AG. .A.C...... A..GGT..T. ....GA.C..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1410       1420       1430       1440       1450
Consensus       WVAVVWWACT CTGRWCCTCA AYCCRCGTGG GGTAAATCCC ACYAACGGYG
TW18            AC.ACAA... ...GT..... .C..G..... .......... ..T.....T.
HL4-9           AA.ACAA... ...GT..... .T..G..... .......... ..T.....T.
BR12            AA.GCAA... ...GT..... .C..G..... .......... ..T.....T.
BZ1             AC.AAAA... ...AA..... .C..A..... .......... ..C.....T.
VK              AA.ACAA... ...GT..... .T..G..... .......... ..T.....T.
QB              AA.ACAA... ...GT..... .T..G..... .......... ..T.....C.
M11             TG.CGTT... ...GA..... .C..G..... .......... ..T.....T.
MX1             TG.CGTT... ...AA..... .C..G..... .......... ..C.....T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1460       1470       1480       1490       1500
Consensus       TTGCCKCGCT TTCDSAAGCG GGTGCAGTTC CTGCRYTGGA GAAGCGTGTY
TW18            .....G.... ...AC..... .......... ....AC.... .........T
HL4-9           .....G.... ...AC..... .......... ....GC.... .........T
BR12            .....T.... ...AG..... .......... ....AT.... .........T
BZ1             .....T.... ...AG..... .......... ....AT.... .........T
VK              .....T.... ...GG..... .......... ....AT.... .........T
QB              .....T.... ...AC..... .......... ....GC.... .........T
M11             .....G.... ...TG..... .......... ....GT.... .........C
MX1             .....G.... ...AG..... .......... ....AT.... .........T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1510       1520       1530       1540       1550
Consensus       ACMRTTTCKG TRTCHCAGCC TTCYCGYAAT CGYAAGAAYT ACAARGTYCA
TW18            ..CG....G. .A..C..... ...T..C... ..C.....C. ....G..C..
HL4-9           ..CG....G. .A..A..... ...T..T... ..T.....C. ....G..C..
BR12            ..CG....G. .A..T..... ...T..T... ..T.....C. ....G..C..
BZ1             ..CA....G. .A..T..... ...T..T... ..T.....C. ....G..T..
VK              ..CG....G. .A..T..... ...C..C... ..T.....C. ....G..C..
QB              ..CG....G. .A..T..... ...T..C... ..T.....C. ....G..C..
M11             ..AA....T. .G..A..... ...T..C... ..T.....T. ....G..T..
MX1             ..AA....T. .A..A..... ...T..C... ..T.....C. ....A..T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       RGTWAARATC CARAACCCRA CCKCTTGYHC TGCAARCGGT WCTTGTGACC
TW18            G..T..G... ..G.....G. ..G....CA. .....A.... T.........
HL4-9           A..T..G... ..G.....G. ..G....TA. .....A.... T.........
BR12            G..T..A... ..G.....G. ..G....TC. .....A.... T.........
BZ1             G..T..A... ..A.....A. ..G....TT. .....A.... T.........
VK              G..T..A... ..A.....G. ..G....TT. .....A.... T.........
QB              G..T..G... ..G.....G. ..G....CA. .....A.... T.........
M11             G..A..G... ..G.....A. ..T....CA. .....G.... A.........
MX1             G..A..G... ..G.....A. ..T....CA. .....G.... A.........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1610       1620       1630       1640       1650
Consensus       CWTCMGTTAC TCGYYMRGCH TATKCTGAYG TGACBTTYTC GTTCACRCAG
TW18            .A..C..... ...CCAG..A ...G....C. ....T..C.. ......G...
HL4-9           .A..C..... ...CCAG..A ...G....C. ....T..T.. ......G...
BR12            .A..C..... ...CCAG..A ...G....C. ....T..T.. ......G...
BZ1             .A..C..... ...CCAG..A ...G....C. ....T..T.. ......G...
VK              .A..C..... ...CCAG..A ...G....C. ....C..T.. ......G...
QB              .A..C..... ...CCAG..A ...G....C. ....C..T.. ......G...
M11             .T..A..... ...TTCA..C ...T....T. ....G..C.. ......A...
MX1             .T..A..... ...TTCG..T ...G....C. ....G..C.. ......G...

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1660       1670       1680       1690       1700
Consensus       TAYAGYACYG WTGAGGAACG HGCWYTYGTW CGHACAGAGC TYVHHGCYCT
TW18            ..T..T..C. A......... A..TT.T..T ..T....... .TATC..T..
HL4-9           ..T..T..C. A......... A..TT.T..T ..T....... .TGTC..T..
BR12            ..T..C..C. A......... A..TT.T..T ..T....... .TGCT..T..
BZ1             ..T..T..C. A......... A..TT.T..T ..T....... .TGCT..T..
VK              ..T..T..C. A......... A..TT.T..T ..C....... .TGCT..T..
QB              ..T..T..C. A......... A..TT.T..T ..T....... .TGCT..T..
M11             ..T..C..T. T......... C..AC.C..A ..A....... .CCAA..T..
MX1             ..C..C..T. A......... T..AC.C..A ..A....... .TAAA..C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1710       1720       1730       1740       1750
Consensus       GYTVGCKRRT CCWMTGYTKR TYRATGCTAT YGAYMRNYTG AAYCCRGCRT
TW18            .C.C..TAG. ..TC..C.GA .CG....... T..TCAGC.. ..T..A..G.
HL4-9           .C.C..TAG. ..TC..T.GA .TG....... C..TCAGC.. ..C..A..G.
BR12            .C.C..TGG. ..TC..C.GA .CG....... C..TCGAC.. ..C..A..G.
BZ1             .C.C..TGA. ..TC..C.GA .CG....... C..TCAGC.. ..C..A..G.
VK              .C.C..TGG. ..TC..C.GA .CG....... C..TCGAC.. ..T..A..A.
QB              .C.C..TAG. ..TC..C.GA .CG....... T..TCAGC.. ..C..A..G.
M11             .T.A..GGA. ..TA..C.TG .CA....... C..CAATC.. ..T..G..G.
MX1             .T.G..GGA. ..AA..C.TA .CG....... C..TAACT.. ..T..G..G.
                  STOP 2
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1760       1770       1780       1790       1800
Consensus       AYTGAACDBY RCTYMTTGCC GGTGDTGGCT CAGGGBMWAR MCCYGDWYCG
TW18            .T.....ACT G..CA..... ....G..... .....TCA.A C..C.ATT..
HL4-9           .T.....ACT G..CA..... ....G..... .....TCT.A A..C.ATC..
BR12            .T.....ATT G..CA..... ....G..... .....GAA.A C..C.ATC..
BZ1             .T.....GTT G..CA..... ....G..... .....GAA.A C..C.ATC..
VK              .T.....ATT G..CA..... ....G..... .....GAA.A C..C.ATC..
QB              .T.....ACT G..CA..... ....G..... .....TCA.A A..C.ATC..
M11             .T.....TGC G..TC...-- -...T..... .....CCT.G C..T.GAC..
MX1             .C.....TGC A..TC...-- -...A..... .....CCT.G C..T.TAC..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1810       1820       1830       1840       1850
Consensus       GKTVHYRWTC CGGAYCCRCC RMYBGAKCCG CCGCCAGGGW CAGGYAVNTA
TW18            .T.ATC---. ....T..A.. GATT..T... .........A ....C.AG..
HL4-9           .T.ATCAT.. ....T..A.. GATC..T... .........A ....T.AA..
BR12            .T.GAT---. ....T..A.. GATT..T... .........A ....T.GT..
BZ1             .T.GAT---. ....T..A.. GATT..T... .........T ....T.GC..
VK              .T.AAT---. ....T..A.. GATT..T... .........A ....T.CC..
QB              .T.ATT---. ....T..A.. GATT..T... .........A ....T.AG..
M11             .G.CCTGA.. ....C..G.. ACCG..G... .........A ....T.GT..
MX1             .G.CCTAA.. ....C..G.. ACTG..G... .........A ....T.GT..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1860       1870       1880       1890       1900
Consensus       TACYTGTCCY TTCSSWATWT GGKMHYTDKM VRRSRTTTAY GARSCTSCKA
TW18            ...C.....C ...GCA..T. ..TCCC.AGA GGAGG....C ..GC..C.T.
HL4-9           ...T.....T ...GCA..T. ..TCCC.AGA GGAGG....C ..GC..C.T.
BR12            ...T.....C ...GCA..T. ..TCTC.GGA AGAGG....T ..GC..C.T.
BZ1             ...C.....T ...GCA..T. ..TCAT.AGA GGAGG....T ..GC..C.T.
VK              ...T.....T ...GCA..T. ..TCTC.AGA GGAGG....T ..AC..C.T.
QB              ...C.....C ...GCA..T. ..TCCC.AGA GGAGG....C ..GC..C.T.
M11             ...C.....T ...CGT..A. ..GATC.GTC CAGCG....T ..AG..G.G.
MX1             ...C.....T ...CGT..A. ..GATC.TTC CAGCA....T ..AG..G.G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1910       1920       1930       1940       1950
Consensus       MYADKDMYCR MYCGTGGSMT ATYTAYHRYG CTRTYGAACT YHVDYCWCGY
TW18            CT.GGAAC.G AC.....CC. ..C..CAAT. ..G.C..... CCGGC.T..C
HL4-9           CT.AGAAC.G AC.....CC. ..C..TAAC. ..A.C..... CCAGC.T..C
BR12            CC.TTGAT.G AC.....CC. ..C..TCGT. ..G.T..... CTCGT.T..C
BZ1             CC.GTGAC.G AC.....CC. ..C..TTAT. ..G.C..... CCAAC.T..T
VK              CT.GTGAC.G AC.....CC. ..C..TCAT. ..G.T..... CCCGC.T..C
QB              CT.AGAAC.G AC.....CC. ..C..TAAT. ..G.T..... CCAGC.T..C
M11             AC.GTTCT.A CT.....GA. ..T..CAAT. ..G.C..... TAGTC.A..C
MX1             AT.GTTCC.A CT.....GA. ..C..TAAC. ..G.C..... CAGTC.T..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1960       1970       1980       1990       2000
Consensus       RADTTTGAYG TYRCCCTYRA WGAYCTYTTG GGYAAYACWR ANTGGCGHGA
TW18            A.G.....T. .TG....CG. A..T..T... ..C..T..AA .T.....C..
HL4-9           A.G.....T. .TG....TA. A..T..T... ..C..T..AA .A.....C..
BR12            A.G.....T. .TG....TG. T..T..C... ..T..C..TA .A.....T..
BZ1             G.T.....T. .TG....TG. T..T..C... ..T..C..TA .A.....A..
VK              G.T.....T. .TG....CG. T..T..C... ..T..C..TG .A.....T..
QB              G.A.....T. .TG....CA. A..T..T... ..C..T..AA .G.....T..
M11             A.T.....C. .CG....TG. T..T..C... ..C..C..AA .C.....T..
MX1             A.A.....C. .TA....TG. T..C..C... ..T..C..AG .T.....C..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2010       2020       2030       2040       2050
Consensus       YTGGGAYKSH MGRCTBAGDT AYACCACGTT YCGCGGTWGC CGWRGYAAYG
TW18            C.....TTCA C.G..G..T. .T........ C......T.. ..TG.C..T.
HL4-9           C.....TTCA C.G..T..T. .T........ T......T.. ..TA.C..C.
BR12            C.....TTCT C.G..T..G. .T........ C......T.. ..TG.C..T.
BZ1             C.....TTCA C.G..T..T. .T........ C......T.. ..TG.C..T.
VK              C.....TTCA C.G..C..A. .T........ C......T.. ..TG.C..C.
QB              T.....TTCT C.G..T..T. .T........ C......T.. ..TG.C..T.
M11             C.....CGGC A.A..G..A. .C........ T......T.. ..TG.C..C.
MX1             T.....CGGC A.A..T..G. .T........ T......A.. ..AG.T..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2060       2070       2080       2090       2100
Consensus       GDTAYATTGA CCTYGAYGCV WCTTMKYTBR YKMMDGAYSA RTRYNWTRCD
TW18            .T..T..... ...C..T..G A...ATC.TG CTACT..TC. G-GCTA.G.G
HL4-9           .T..T..... ...T..T..G T...ATC.TG CTACT..TC. G-GCTA.G.G
BR12            .T..T..... ...T..T..G A...ATC.CG CTACT..TC. A-GCTA.G.T
BZ1             .T..T..... ...T..T..G A...ATC.TG CTACT..CC. A-GCGA.G.T
VK              .T..T..... ...T..T..G A...ATC.CG CTACT..CC. A-GCAA.G.T
QB              .T..T..... ...T..T..G A...ATC.TG CTACT..TC. G-GCTA.G.G
M11             .G..C..... ...T..C..A A...CGT.GA TGAAA..TG. A.ATCT.A.A
MX1             .A..C..... ...T..C..C A...CGT.GA TGCAG..TG. A.ATCT.A.A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2110       2120       2130       2140       2150
Consensus       YSDTCAGAAG TAYBWWRTKC GYRMRGGSAA RMRWCCYGGT GYYTTYGGYW
TW18            TGA....... ..CGATA.T. .CGAG..C.. AAAA..T... .CT..C..TA
HL4-9           TGA....... ..CGATA.T. .CGAA..C.. GAAA..C... .CT..C..TA
BR12            CGA....... ..CGATA.T. .CACG..C.. AAGA..T... .CC..T..CA
BZ1             TGA....... ..CGATA.T. .CACG..C.. GAAA..C... .CC..C..TA
VK              TGA....... ..CGATA.T. .CACA..C.. AAGA..T... .CC..C..TA
QB              TGA....... ..TGATA.T. .CGAG..C.. GAAA..T... .CT..C..TA
M11             TCG..-.... ..TCTTG.G. .TGAA..G.. ACGT..C... .TC..C..TA
MX1             TCT..-.... ..TTTAG.G. .TGAG..G.. ACGT..C... .CT..C..TT


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2160       2170       2180       2190       2200
Consensus       MHRTYGARCR RTTYRTYTAT CTTAARTCGA TAAAYGCHTA YTGYTCKCTY
TW18            ACA.T..G.G A..CA.T... .....G.... ....T..T.. T..C..T..T
HL4-9           ACG.T..G.G A..CA.T... .....G.... ....T..T.. T..C..T..T
BR12            AAA.T..A.A A..TA.T... .....G.... ....T..A.. T..T..T..C
BZ1             ACA.C..A.A A..CA.C... .....G.... ....T..C.. T..T..T..T
VK              ATA.T..A.A A..TA.T... .....G.... ....T..A.. T..T..T..T
QB              ACA.T..G.G A..CA.T... .....G.... ....T..T.. T..C..T..T
M11             ACA.C..G.G G..TG.T... .....G.... ....C..T.. C..C..G..C
MX1             CAA.T..A.G G..TG.T... .....A.... ....C..A.. C..T..T..T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2210       2220       2230       2240       2250
Consensus       AGCGATATTR CGGCCTAYCR CDCCGAYGGH GTRRTAGTTG GCTTTTGGCG
TW18            .........G .......T.A .G....T..T ..GA...... ..........
HL4-9           .........G .......T.A .G....T..T ..GA...... ..........
BR12            .........G .......T.A .G....C..C ..AA...... ..........
BZ1             .........G .......T.A .G....T..T ..GA...... ..........
VK              .........G .......T.A .G....T..C ..AA...... ..........
QB              .........G .......T.A .G....T..C ..GA...... ..........
M11             .........A .......T.G .A....T..T ..GA...... ..........
MX1             .........A .......C.A .T....C..A ..GG...... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2260       2270       2280       2290       2300
Consensus       YGAYCCRTCH AGYGGKGGWG CCATACCRTT YGACTTYAVN VARTTTGAYW
TW18            C..T..A..T ..T..T..T. .......G.. T.....C.CT A.G.....TA
HL4-9           C..T..A..T ..T..T..T. .......G.. T.....C.CT A.G.....TA
BR12            C..T..A..T ..T..T..T. .......G.. T.....C.CT C.G.....TA
BZ1             C..T..A..T ..T..T..T. .......G.. T.....C.CA G.G.....TA
VK              C..T..A..T ..T..T..T. .......G.. T.....C.CG G.A.....TA
QB              C..T..A..C ..T..T..T. .......G.. T.....C.CT A.G.....TA
M11             T..C..G..A ..C..G..A. .......A.. C.....T.AC G.A.....CT
MX1             C..C..G..A ..C..G..A. .......A.. C.....T.GC G.G.....TT
                                                              STOP 3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2310       2320       2330       2340       2350
Consensus       MGAMYAAATG TCCDATHCAA GCYGTKATWG TCGTTCCTCG TSYTTAGTAA
TW18            A..CT..... ...T..T... ..C..G..A. .......... .GC.......
HL4-9           A..CT..... ...T..T... ..C..G..A. .......... .GC.......
BR12            A..CT..... ...T..T... ..C..G..A. .......... .GC.......
BZ1             A..CT..... ...A..T... ..T..G..A. .......... .GC.......
VK              A..CT..... ...G..T... ..C..G..A. .......... .GC.......
QB              A..CT..... ...T..T... ..C..G..A. .......... .GC.......
M11             C..AC..... ...G..A... ..C..T..A. .......... .CT.......
MX1             C..AC..... ...G..C... ..C..T..T. .......... .CT.......
                           ORF4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2360       2370       2380       2390       2400
Consensus       CTRARGRWGA WMTGCATGTC WAAGACABYR YMBTCGCRTA ASTCKCTYAG
TW18            ..A.G.AT.. AA........ T......GCA TCT....A.. .C..T..T..
HL4-9           ..A.G.AT.. AA........ T......GCA TCT....A.. .C..T..T..
BR12            ..A.G.AT.. AA........ T......GCA TCT....A.. .C..T..T..
BZ1             ..A.G.AT.. AA........ T......GCA TCC....G.. .C..T..C..
VK              ..A.G.AT.. AA........ T......GCA TCT....A.. .C..T..T..
QB              ..A.G.AT.. AA........ T......GCA TCT....G.. .C..T..C..
M11             ..G.A.GA.. TC........ T......TCG CAG....G.. .G..G..T..
MX1             ..A.A.GA.. TC........ A......CTG CAG....G.. .G..G..T..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2410       2420       2430       2440       2450
Consensus       CGSAMAAYTS CGCCGHGCYG CGAACACAAG AATYGWGGTH GAAGRTAACC
TW18            ..C.C..T.G .....A..T. .......... ...T.A...A ....G.....
HL4-9           ..C.C..T.G .....A..T. .......... ...T.A...C ....G.....
BR12            ..C.C..T.G .....A..T. .......... ...T.A...T ....G.....
BZ1             ..C.C..T.G .....A..C. .......... ...T.A...C ....G.....
VK              ..C.C..T.G .....A..T. .......... ...T.A...C ....G.....
QB              ..C.C..T.G .....A..C. .......... ...T.A...T ....G.....
M11             ..G.A..C.C .....T..T. .......... ...C.T...T ....A.....
MX1             ..G.A..C.C .....C..T. .......... ...C.T...T ....G.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2460       2470       2480       2490       2500
Consensus       TCGCACTKTC YATYGCVAAY GAYYTAHTRT YBGCNYWDGR TSWDKMRYCR
TW18            .......T.. C..T..C..C ..TT..C.A. TG..TTAT.G .CAATCGC.A
HL4-9           .......T.. C..C..C..C ..TC..C.G. TG..TTAT.G .CAATCGC.A
BR12            .......T.. C..T..C..C ..CT..A.G. TG..ATAT.G .CAATCGC.A
BZ1             .......T.. C..T..C..C ..TT..A.G. TG..TTAT.G .CAATCGC.A
VK              .......G.. C..T..C..C ..TT..A.G. TG..ATAT.G .CAATCGC.A
QB              .......T.. C..T..C..C ..TT..C.G. TG..CTAT.G .CAGTCGC.A
M11             .......T.. C..C..A..C ..TC..T.G. CT..GTTG.A .GTTGAGT.G
MX1             .......G.. T..C..G..T ..TC..T.A. CC..GCTA.A .GTAGAAC.G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2510       2520       2530       2540       2550
Consensus       TTTARYTCBG ARKMNGASTG TATHWSHYKY DSHCCGARAT TYGRCVDVWC
TW18            ....AC..T. .AGCT..G.. ...TTCACTC AGT....G.. .C.A.GGGA.
HL4-9           ....AC..C. .AGCT..G.. ...TTCACTC AGT....G.. .T.A.GGGA.
BR12            ....GT..T. .GTCT..G.. ...TTCACTC AGC....A.. .C.A.GGAA.
BZ1             ....AT..C. .GTCC..G.. ...TTCACTC GGT....A.. .C.A.GAAA.
VK              ....GT..C. .GTCT..G.. ...TTCACTC AGC....A.. .C.A.GGAA.
QB              ....AC..T. .GGCT..G.. ...TTCATTC AGC....G.. .C.A.GGGA.
M11             ....GT..G. .GGAG..C.. ...CAGCCGC TCC....A.. .C.A.CTCT.
MX1             ....AT..G. .AGAA..C.. ...AAGTCGT TCA....A.. .C.G.ATCT.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2560       2570       2580       2590       2600
Consensus       HSCRGAYVAM TTTAGGAWWW MYTATCTTVV WGCYGAGRTH ATGTCGAAGT
TW18            CC.G..TG.C .......TAA AT......AA A..C...G.C ..........
HL4-9           CC.G..TC.C .......TAA AT......AA A..C...G.C ..........
BR12            CC.G..TA.C .......TAA AT......AA A..C...A.C ..........
BZ1             CC.G..CA.C .......TAA AT......AA A..C...A.C ..........
VK              CC.G..TA.C .......TAA AT......AA A..C...A.C ..........
QB              CC.G..TG.C .......TAA AT......AA A..C...A.C ..........
M11             TG.G..TC.A .......ATT CC......GC T..C...A.T ..........
MX1             AC.A..TC.A .......ATT CC......CG T..T...A.A ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2610       2620       2630       2640       2650
Consensus       AYGAYKMHTT YAGCCTAGGT ATYRATACCG AAGCYGYWGC MTGGRARAAG
TW18            .T..CGAT.. C......... ..TG...... ....T.CT.. C...G.A...
HL4-9           .T..CGAT.. C......... ..TG...... ....T.CT.. C...G.A...
BR12            .T..CGAT.. C......... ..TG...... ....T.CT.. A...A.G...
BZ1             .T..CGAT.. C......... ..TG...... ....T.CT.. A...G.G...
VK              .T..CGAT.. C......... ..CG...... ....T.CT.. A...G.A...
QB              .T..CGAC.. C......... ..TG...... ....T.TT.. C...G.G...
M11             .T..CTCA.. C......... ..TA...... ....C.TA.. C...G.G...
MX1             .C..TTCT.. T......... ..TA...... ....C.TA.. A...G.A...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2660       2670       2680       2690       2700
Consensus       TTYCTRGCWG CRGAGGCTGA RTGTGCTWWR ACGAAYSHKM GHCTCTATAG
TW18            ..C..G..A. .A........ A......TTA .....CGCTC .T........
HL4-9           ..C..G..A. .A........ A......TTA .....CGCTC .T........
BR12            ..T..A..A. .A........ A......TTA .....CGCTC .T........
BZ1             ..C..A..A. .A........ A......TTA .....CGCTC .T........
VK              ..C..A..A. .A........ A......TTA .....CGCTC .C........
QB              ..C..G..A. .A........ A......TTA .....CGCTC .T........
M11             ..C..G..T. .G........ G......ATA .....TCAGC .T........
MX1             ..C..A..T. .G........ G......AAG .....TCTGA .A........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2710       2720       2730       2740       2750
Consensus       RCCTRACTAC ARTGAGGATT TCAATTTCTC AYTGGGYGAG DCRTGTMTWC
TW18            G...G..... .G........ .......... .C....C... T.A...A.T.
HL4-9           G...G..... .G........ .......... .C....C... T.A...A.T.
BR12            A...A..... .G........ .......... .C....T... T.G...C.T.
BZ1             G...A..... .G........ .......... .C....C... T.A...A.T.
VK              A...A..... .G........ .......... .C....T... T.A...C.T.
QB              G...G..... .G........ .......... .C....C... T.A...A.A.
M11             A...A..... .A........ .......... .T....T... G.A...A.T.
MX1             G...A..... .A........ .......... .T....T... A.A...A.T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2760       2770       2780       2790       2800
Consensus       ACATGGCYCG YMGAAAAATA GYYAAGYTRW TAGGAGAYKY NSYKYCSKTT
TW18            .......T.. TA........ .CC...T.AA .......CGC TCCGT.CG..
HL4-9           .......T.. CA........ .TC...C.AA .......TGC TCCGT.CG..
BR12            .......T.. CC........ .TT...T.GA .......TGC TCCGT.CG..
BZ1             .......T.. TA........ .TT...C.AA .......CGC GCCGT.CG..
VK              .......T.. CC........ .TT...T.AA .......CGC GCCGT.CG..
QB              .......T.. TA........ .CC...C.AA .......TGT TCCGT.CG..
M11             .......C.. CC........ .TT...T.AT .......CTC CGTTC.GT..
MX1             .......T.. CC........ .TT...C.AT .......TTC AGTTC.GT..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2810       2820       2830       2840       2850
Consensus       GAGGSTRTGY TGCGHCAYTG CCGDTTTTCY GGYGGTGCYA CAACRACGAA
TW18            ....G.A..T ....T..C.. ...T.....C ..T.....C. ....A.....
HL4-9           ....G.A..T ....T..C.. ...A.....T ..C.....C. ....A.....
BR12            ....G.A..T ....T..T.. ...A.....T ..C.....C. ....A.....
BZ1             ....G.A..C ....T..C.. ...T.....C ..T.....T. ....A.....
VK              ....G.A..T ....C..T.. ...G.....T ..C.....C. ....A.....
QB              ....G.A..T ....T..C.. ...A.....T ..C.....T. ....A.....
M11             ....C.A..T ....A..C.. ...T.....T ..T.....C. ....G.....
MX1             ....C.G..T ....A..T.. ...G.....C ..C.....C. ....A.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2860       2870       2880       2890       2900
Consensus       TARCCGTYYR YAYGGYCATC CGTCCTTCAA GTTTGCKCTT VCRCAAGMGT
TW18            ..A....TCG T.C..T.... .......... ......G... C.A....C..
HL4-9           ..A....TCG C.C..C.... .......... ......G... C.A....C..
BR12            ..A....TCA T.T..T.... .......... ......G... C.A....C..
BZ1             ..A....TCA T.C..T.... .......... ......G... C.A....C..
VK              ..A....TCA T.C..T.... .......... ......G... C.A....C..
QB              ..A....TCG T.C..T.... .......... ......G... C.G....C..
M11             ..A....TCA T.C..T.... .......... ......T... A.G....A..
MX1             ..G....CTA T.C..C.... .......... ......T... G.G....A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2910       2920       2930       2940       2950
Consensus       GTACSCCWCG GGCTKTKMMR TAYGTKYWRG CYYTHARRGC YTSWACRVRY
TW18            ....G..T.. ....T.GAAA ..T..TTTA. .TC.T.GA.. T.CT..ACAT
HL4-9           ....G..T.. ....T.GAAA ..T..TTTG. .TC.T.GA.. T.CT..ACAT
BR12            ....G..T.. ....T.GAAA ..T..TTTG. .TC.T.GA.. T.CA..GCAT
BZ1             ....G..T.. ....T.GAAG ..T..TCTG. .TC.T.GA.. C.CT..GCAC
VK              ....G..T.. ....T.GAAA ..C..TTTG. .TC.C.GA.. C.CA..ACAT
QB              ....G..T.. ....T.GAAG ..T..TTTA. .TC.C.GA.. T.CT..ACAT
M11             ....G..A.. ....G.TCCA ..C..GCAG. .CC.T.AG.. C.GT..AGGT
MX1             ....C..A.. ....G.TCCA ..T..GCAA. .TT.A.AG.. C.GT..AAAC


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2960       2970       2980       2990       3000
Consensus       WTSGAYMTYR GDRTYWCYRA KRTYAGCCCT TTYAATAAAG CAGTTACTGT
TW18            T.C..TA.CA .AA.TT.TG. TA.T...... ..T....... ..........
HL4-9           T.C..TA.CA .GG.TT.TG. TA.T...... ..T....... ..........
BR12            T.C..TA.CA .AG.TT.TG. TA.T...... ..T....... ..........
BZ1             T.C..TA.CA .AG.TT.TG. TA.T...... ..T....... ..........
VK              T.C..TA.CA .AG.TT.TG. TA.T...... ..T....... ..........
QB              T.C..TA.CA .AA.TT.TG. TA.T...... ..T....... ..........
M11             A.G..CC.TG .TA.CA.CA. GG.C...... ..C....... ..........
MX1             A.G..CC.TG .TA.TA.CA. GG.T...... ..C....... ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3010       3020       3030       3040       3050
Consensus       ACCWAAGAAC AGTAARACHG ATCGYTGYAT HGCYATCGAR CCHGGYTGGA
TW18            ...T..... .....A..C. ....T..T.. T..T.....G ..T..C....
HL4-9           ...T..... .....A..C. ....T..T.. T..T.....A ..C..C....
BR12            ...A..... .....A..C. ....C..T.. T..T.....A ..C..T....
BZ1             ...A..... .....A..C. ....C..T.. C..T.....A ..C..T....
VK              ...A..... .....G..C. ....C..T.. T..T.....A ..C..T....
QB              ...T..... .....G..A. ....T..T.. T..T.....A ..T..T....
M11             ...A..... .....A..T. ....C..T.. A..C.....A ..T..C....
MX1             ...A..... .....A..T. ....C..C.. C..T.....G ..A..C....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3060       3070       3080       3090       3100
Consensus       ATATGTTTTT YCARYTRGGY ATYGGTGGYR THHTWCGYGA HMRGYTGCRY
TW18            .......... T..AC.A..C ..C.....CA .TC.A..C.. TCG.T...GT
HL4-9           .......... T..AC.A..C ..C.....CA .CC.A..C.. TCG.T...GT
BR12            .......... C..AT.A..T ..C.....CA .TT.A..T.. TCG.T...GT
BZ1             .......... T..AC.A..T ..C.....CA .CT.A..C.. CCG.C...GT
VK              .......... C..AT.A..T ..C.....CA .CC.A..C.. TCG.T...GT
QB              .......... C..AC.G..T ..C.....CA .TC.A..C.. TCG.T...GT
M11             .......... C..GC.A..T ..C.....TG .TA.A..C.. AAA.T...GT
MX1             .......... C..GT.G..C ..T.....CG .AA.T..C.. AAA.T...AC


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3110       3120       3130       3140       3150
Consensus       TKBTGGRRYA TCGAYCTGAA YGATCAGACG ATWAAYCARV NBCGCGCWYA
TW18            .GC...GGT. ....T..... T......... ..A..T..GC GC.....TC.
HL4-9           .GC...GGT. ....T..... C......... ..A..T..GC AT.....TC.
BR12            .GT...GGT. ....C..... T......... ..A..C..GC AC.....TC.
BZ1             .GC...GGT. ....T..... T......... ..A..C..GC GT.....TC.
VK              .GC...GGT. ....C..... T......... ..A..C..GC GC.....TC.
QB              .GC...GGT. ....T..... T......... ..A..T..GC GC.....TC.
M11             .TG...GGC. ....T..... T......... ..T..C..AA CG.....AT.
MX1             .TG...AAT. ....C..... T......... ..T..C..GG TG.....AT.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3160       3170       3180       3190       3200
Consensus       YKHRGGCWSY BKTRVYRATR AHYTHGCNAC RGTKGATCTC TCRRSVGCDA
TW18            CGAA...TCC GT.ACTA..A .TT.A..A.. G..T...... ..AGCG..A.
HL4-9           CGAA...TCC GT.ACTA..A .TT.A..A.. G..T...... ..AGCG..A.
BR12            CGAG...TCC GT.ACTA..G .TT.A..A.. G..T...... ..AGCG..A.
BZ1             CGAA...TCT GT.ACTA..A .CT.A..A.. A..T...... ..AGCG..A.
VK              CGAA...TCC GT.ACTA..G .TT.A..G.. G..T...... ..AGCA..A.
QB              CGAA...TCC GT.ACTA..A .CT.A..A.. G..T...... ..AGCG..A.
M11             TTTA...AGC CG.GATG..A .TC.C..C.. G..G...... ..AAGA..T.
MX1             TTCA...AGC TG.AGCA..G .AC.T..T.. A..G...... ..GAGC..G.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3210       3220       3230       3240       3250
Consensus       GYGATWCTAT WTCKCTTGCB CTYKKYGAGC TCYTDHTRCC YCCWGVBTGG
TW18            .C...T.... A..T.....T ..CTGT.... ..T.AC.G.. C..A.GC...
HL4-9           .C...T.... A..T.....T ..CTGT.... ..T.AT.G.. C..A.GT...
BR12            .C...T.... A..T.....T ..CTGT.... ..T.AT.A.. T..A.GT...
BZ1             .C...T.... A..T.....T ..CTGC.... ..T.GT.G.. T..A.GG...
VK              .C...T.... A..T.....T ..CTGC.... ..T.AT.G.. T..A.GT...
QB              .C...T.... A..T.....C ..CTGT.... ..T.AT.G.. C..A.GC...
M11             .C...A.... T..G.....C ..TGTT.... ..C.TA.G.. T..T.AG...
MX1             .T...A.... T..G.....G ..CGTT.... ..C.GC.A.. C..T.CG...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3260       3270       3280       3290       3300
Consensus       TTTRARGTYC TKWYGGMYYT HMGRTCACSH ARRGGBMDVT TGCCWRAYGG
TW18            ...G.G..C. .TAT..ACC. CA.A....CT .AA..GCGG. ....TG.C..
HL4-9           ...G.G..T. .TAT..ATC. CA.A....CA .AA..GCGG. ....TG.C..
BR12            ...G.G..C. .TAC..ATC. CA.A....CT .AG..GCAA. ....TG.C..
BZ1             ...G.A..C. .TAT..ACC. TC.A....CT .AG..GCGA. ....TA.C..
VK              ...G.G..C. .TAT..ACC. CA.A....CT .AG..GCGA. ....TG.C..
QB              ...G.G..T. .TAT..ACC. CA.A....CT .AG..GCGA. ....TG.C..
M11             ...A.G..C. .GTT..CCT. AA.A....CC .AG..CATC. ....AG.T..
MX1             ...A.A..C. .TAC..ACC. TA.G....GA .GG..TATG. ....AG.C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3310       3320       3330       3340       3350
Consensus       TASHRTYRTT AYYTAYGAGA ARATWTCYTC NATGGGTAAY GGHTWYACMT
TW18            ..GTG.TG.. .CC..T.... .G..T..C.. T........C ..A.AC..A.
HL4-9           ..GTG.TG.. .CC..T.... .A..A..C.. G........C ..A.AC..A.
BR12            ..GTG.TA.. .CC..T.... .A..T..T.. C........T ..T.AC..A.
BZ1             ..GTG.CG.. .CT..T.... .G..T..T.. C........C ..T.AC..A.
VK              ..GTG.CA.. .TC..C.... .G..T..T.. C........T ..T.AC..A.
QB              ..GTG.TG.. .CC..C.... .G..T..T.. T........C ..T.AC..A.
M11             ..CCG.CA.. .CT..T.... .A..A..C.. A........T ..C.AT..C.
MX1             ..GAA.CA.. .CC..T.... .A..T..C.. A........C ..T.TC..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3360       3370       3380       3390       3400
Consensus       TCGARCTYGA GTCGCTTATH TTYGCDKCTC TYGCYCGKTC YKTWTGYGAR
TW18            ....G..T.. .........C ..T..TT... .T..T..T.. CG.T..T..G
HL4-9           ....G..T.. .........C ..T..TT... .T..T..T.. CG.T..T..G
BR12            ....G..C.. .........T ..T..TT... .C..C..T.. CG.T..T..G
BZ1             ....A..C.. .........T ..T..TT... .T..T..T.. CG.T..T..G
VK              ....G..T.. .........C ..C..TT... .T..T..T.. CG.T..T..G
QB              ....G..C.. .........T ..T..TT... .C..T..T.. CG.T..T..G
M11             ....G..T.. .........A ..T..GG... .T..T..G.. TT.A..C..A
MX1             ....G..C.. .........A ..T..AG... .T..T..G.. TT.A..C..G
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3410       3420       3430       3440       3450
Consensus       WTACTGRRCT TASRMYCRTC DRRKGTYACK GTYTAYGGMG AYGAYATWAT
TW18            A.....GA.. ..GACT.A.. TGAG..C..T ..C..C..A. .C..T..T..
HL4-9           A.....GA.. ..GACT.G.. TGAG..C..T ..T..C..A. .C..T..T..
BR12            A.....GA.. ..GACT.A.. TGAG..C..T ..T..C..A. .T..C..T..
BZ1             A.....GA.. ..GACT.A.. TGAG..C..T ..C..C..A. .C..C..T..
VK              A.....AA.. ..GACT.A.. TGAG..C..T ..T..C..A. .T..C..T..
QB              A.....GA.. ..GACT.G.. TGAG..C..T ..T..C..A. .C..T..T..
M11             T.....GG.. ..CGAC.G.. AGAT..T..G ..C..T..C. .T..C..A..
MX1             T.....AA.. ..CAAC.G.. GAGT..C..G ..C..T..C. .T..T..T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3460       3470       3480       3490       3500
Consensus       HTTRCCRTCM BRYGCRKKYM SYBCBYTNVD KGAAGTYTTY WMSTAYGTWG
TW18            C..A..G..C TGT..AGTCC CTG.CC.TCG G.....C..T AAG..T..T.
HL4-9           T..A..G..C TGT..AGTCC CTG.CC.TCG G.....T..T AAG..T..T.
BR12            T..G..A..C CGT..AGTCC CTG.TC.TCA G.....C..T AAG..T..T.
BZ1             T..G..G..C TGT..AGTCC CTG.TC.TCA G.....C..C AAG..T..T.
VK              T..G..A..C CGT..AGTTC CTG.TC.TCA G.....C..T AAG..T..T.
QB              T..A..G..C TGT..AGTCC CTG.CC.CCG G.....T..T AAG..T..T.
M11             A..G..A..A GAC..GTGCA GTC.TC.AGT T.....T..C TCC..T..T.
MX1             A..G..A..A GAC..GTGCA GCT.GT.GAT T.....T..C TCC..C..A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3510       3520       3530       3540       3550
Consensus       GTTTTMSDAC CAAYRMKAAR AARACBTTYT YYRRNGGRCC GTTCMGAGAG
TW18            .....ACA.. ...TACT..A ..G..T..T. CCGAG..G.. ....A.....
HL4-9           .....ACG.. ...TACT..A ..G..T..C. CCGAG..G.. ....A.....
BR12            .....ACT.. ...TACT..A ..G..T..C. CTGAG..G.. ....A.....
BZ1             .....ACT.. ...TACT..A ..G..C..T. CCGAA..G.. ....A.....
VK              .....ACT.. ...TACT..G ..G..T..T. CTGAG..G.. ....A.....
QB              .....ACG.. ...TACT..A ..G..T..T. CCGAG..G.. ....A.....
M11             .....CGT.. ...CAAG..G ..A..G..T. CTAGT..A.. ....C.....
MX1             .....AGA.. ...CGAG..G ..G..C..T. TCGAC..G.. ....C.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3560       3570       3580       3590       3600
Consensus       TCGTGCGGHA AGCACTACTW TWYKGGCGTW GAYGTYACWC CYTTYTACAT
TW18            ........T. .........A .TCT.....A ..T..T..T. .C..C.....
HL4-9           ........T. .........A .TCT.....A ..T..T..T. .C..T.....
BR12            ........T. .........A .TCT.....A ..T..T..T. .C..T.....
BZ1             ........T. .........A .TCT.....A ..T..T..T. .C..T.....
VK              ........T. .........A .TCT.....A ..T..T..T. .C..T.....
QB              ........C. .........A .TCT.....A ..T..T..T. .C..T.....
M11             ........A. .........T .TTG.....T ..C..C..A. .T..C.....
MX1             ........A. .........T .ATG.....T ..C..C..A. .T..C.....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3610       3620       3630       3640       3650
Consensus       ACGYCRCCGY ATAGTGAVYC CYDCYGATYT MATAYTGGTT TTGAAYMASM
TW18            ...T.A...T .......CC. .TG.C...T. A...C..... .....TA.CC
HL4-9           ...T.A...T .......AT. .TA.C...T. A...T..... .....TA.CC
BR12            ...T.G...C .......GT. .TG.C...T. A...T..... .....TA.CC
BZ1             ...C.G...C .......GT. .TG.C...T. A...T..... .....TA.CC
VK              ...T.G...T .......GT. .TG.C...T. A...T..... .....TA.CC
QB              ...T.A...T .......GT. .TG.C...T. A...C..... .....TA.CC
M11             ...T.G...T .......GT. .CT.C...C. C...C..... .....CC.GA
MX1             ...C.A...T .......GT. .CT.T...C. C...C..... .....CC.GA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3660       3670       3680       3690       3700
Consensus       TRTATCGKTG GGCCACRATT GAYGGCGTAT GGGATCCTAG RGYMYATYCY
TW18            .A.....G.. ......G... ..T....... .......... G.CCC..C.C
HL4-9           .A.....G.. ......G... ..T....... .......... G.CCC..T.C
BR12            .A.....G.. ......G... ..C....... .......... A.CCC..T.T
BZ1             .A.....G.. ......G... ..C....... .......... G.CCC..T.C
VK              .A.....G.. ......G... ..C....... .......... A.CAC..T.C
QB              .A.....G.. ......A... ..C....... .......... G.CCC..T.T
M11             .G.....T.. ......A... ..C....... .......... G.TAT..C.T
MX1             .G.....T.. ......G... ..T....... .......... G.TAT..C.T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3710       3720       3730       3740       3750
Consensus       GTRTAYMYCA AGTAYMGWMR KYWSCTKCCK RAWMWKCTSC RVMGDAATRY
TW18            ..G..CCT.. ....TC.TAA GTTG..G..T A.ACAG..G. AAC.T...AC
HL4-9           ..A..CCT.. ....TC.TAA GTTG..G..T A.ACAG..G. AAC.T...AC
BR12            ..A..TCT.. ....TC.TAA GCTG..G..T A.ACAG..G. AGC.T...AC
BZ1             ..G..CCT.. ....TC.TAA GTTG..G..T A.ACAG..G. AGC.T...AC
VK              ..A..CCT.. ....TC.TAA GCTG..G..T A.ACAG..G. AGC.T...AC
QB              ..G..CCT.. ....TC.TAA GTTG..G..T A.ACAG..G. AAC.T...AC
M11             ..A..CAC.. ....TA.ACG TTAC..T..G G.AATT..C. GGA.G...GT
MX1             ..A..CAC.. ....CA.ACG TCTG..G..G G.TATT..C. GCA.A...GT


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3760       3770       3780       3790       3800
Consensus       YRTRCCTGAT GGWTAYGGTG ATGGNGCYCT CGTCGGATCK GTCYTDAYCA
TW18            TA.A...... ..T..C.... ....G..C.. .........G ...C.G.T..
HL4-9           TA.A...... ..T..C.... ....A..C.. .........G ...T.G.T..
BR12            TA.A...... ..T..C.... ....T..T.. .........G ...T.G.T..
BZ1             TA.A...... ..T..C.... ....C..C.. .........G ...C.G.T..
VK              TA.A...... ..T..C.... ....C..T.. .........G ...T.G.T..
QB              TA.A...... ..T..C.... ....T..C.. .........G ...C.A.T..
M11             CG.G...... ..A..C.... ....T..C.. .........T ...T.A.T..
MX1             CG.A...... ..A..T.... ....T..C.. .........T ...C.T.C..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3810       3820       3830       3840       3850
Consensus       RTCCHTTCGC RRAAAAYCGC GGDTGGRTYC GGYRYGTDCC GRTGATHAYD
TW18            A...T..... GA....C... ..G...A.C. ..TAC..A.. .G....C.CG
HL4-9           A...T..... GA....C... ..G...A.C. ..TAC..A.. .G....C.TA
BR12            A...T..... GA....C... ..G...A.C. ..TAT..T.. .G....C.TA
BZ1             A...A..... GA....C... ..G...A.C. ..TAT..A.. .G....C.TT
VK              A...C..... GA....C... ..A...A.C. ..TAT..T.. .G....C.TA
QB              A...T..... GA....C... ..G...A.C. ..TAC..A.. .G....T.CG
M11             G...T..... AG....T... ..T...G.T. ..CGT..G.. .A....T.TA
MX1             G...T..... AG....T... ..T...G.T. ..CGT..G.. .A....A.TT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3860       3870       3880       3890       3900
Consensus       GACMABAVRA RRGACCGAGW DCGYRHYGAR YNDGGDTCRT ATCTYTAYGA
TW18            ...C.T.CA. GG.......A G..CGCT..G TTA..G..G. ....T..C..
HL4-9           ...C.C.CA. GG.......A G..CGTT..A TCG..G..G. ....T..C..
BR12            ...C.T.CA. GG.......A A..CACT..G TCA..A..G. ....C..C..
BZ1             ...C.C.CA. GG.......A G..CACT..G CCG..G..A. ....C..T..
VK              ...C.T.CA. GG.......A A..CACC..G TCG..A..G. ....C..C..
QB              ...C.T.CA. GG.......A G..CGCT..G TTG..G..G. ....C..C..
M11             ...A.G.GG. AA.......T T..TGAC..A TAT..T..G. ....C..C..
MX1             ...A.G.AG. AA.......T G..TGAC..G CGT..T..A. ....C..T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3910       3920       3930       3940       3950
Consensus       NCTHTKBTCK YKNYRBYWVY YSGAADGTRA CRRTGRGTTS CCBYTHARSG
TW18            C..C.TC..G CGTTGTCTCC CG...A..A. .GA..G...G ..TC.A.GG.
HL4-9           C..C.TT..G CGTTGTTTCT CG...G..A. .GA..G...G ..TT.A.GG.
BR12            T..T.TC..G CGCTGCTTCT CG...A..A. .GA..G...G ..TC.A.GG.
BZ1             T..C.TC..T CGCTGTTTCT CG...A..A. .GA..G...G ..CC.A.GG.
VK              T..T.TC..G CGCTATTTCT CG...A..A. .GA..G...G ..TC.A.GG.
QB              C..C.TC..G CGTTGTCTCT CG...A..A. .GA..G...G ..TC.T.GG.
M11             G..A.GG..G TTGCAGCAAC TC...T..G. .AG..A...C ..CT.T.AC.
MX1             A..A.GG..G TTACAGCAGC TC...T..G. .AG..A...C ..GT.C.AC.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3960       3970       3980       3990       4000
Consensus       GKYCRYYGRK YKKYGRTHCY RHNDATSBVN YYBBYRYHKA YSVRCWTATH
TW18            .TC.GTC.AG TTGC.A.T.T GTAA..CTGT CTGCTGTCG. TCAA.T...C
HL4-9           .TC.GTC.AG TTGC.A.T.T GTAG..CTAT CTGCCATTG. TCAA.T...C
BR12            .TC.GTC.AG TTGC.A.C.T GTAT..CCAC TTGCCATAG. TCAG.T...A
BZ1             .TC.GTC.AG TTGC.A.T.T ATAT..CGGT TTGCCGTTG. CCAA.T...A
VK              .TC.GTC.AG TTGC.A.T.T ATAT..CCGC TCGCTATCG. TCAG.T...T
QB              .TC.ATC.GG TTGC.A.T.T GCGG..CTAT TTGCCATCG. TCAG.T...C
M11             .GT.GCT.GT CGTT.G.T.C ACTG..GGCA CTCTCGCTT. CGCA.A---C
MX1             .GT.GCT.GT TGTT.G.A.C AACG..GGAG TTTGCACTT. CCGA.A---C
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4010       4020       4030       4040       4050
Consensus       YGWRRRMGBD WWYCTACVRH DRTMAGYRRD KCBRYHRGBR MRTWYGAYRT
TW18            T.TAGGA.CA ATC....AAA GA.A..TAGA T.TACTG.CA AG.TC..TA.
HL4-9           T.TAGGA.TA ATC....AAA GA.A..TAGA T.TACCG.TA AG.TC..TA.
BR12            T.TAAGA.TA ATC....AAA AA.A..CAGA T.TACTG.TA AG.TT..TG.
BZ1             T.TAAGA.CA ATC....AAA AG.A..CAGA T.TACTG.GA AA.TT..TA.
VK              T.TAAGA.TA ATC....AAA AA.A..CAGA T.TACTG.TA AG.TC..TA.
QB              T.TAGGA.TA ATC....GAA GA.A..CAGG T.TACCG.CA AA.TC..TA.
M11             C.AGAAC.GT TAC....CGT TA.C..TGAT G.CGTAA.TG CG.TT..CA.
MX1             C.AGAGC.GG TTT....CGC TA.C..TGAT T.GGTTG.CG CA.AC..CA.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4060       4070       4080       4090       4100
Consensus       MVWRTRKATM SCGTGCRGTA GYCGTGTYCT GGCWCCCTAC GGGGWYTTCC
TW18            ACAG.AT..C G.....A... .C.....T.. ...A...... ....TC....
HL4-9           ACAG.AT..C G.....G... .C.....T.. ...A...... ....TT....
BR12            ACAG.AT..C G.....A... .C.....T.. ...A...... ....TC....
BZ1             ACAG.AT..C G.....A... .C.....T.. ...A...... ....TT....
VK              ACAG.AT..C G.....A... .C.....T.. ...A...... ....TT....
QB              ACAG.AT..C G.....A... .C.....T.. ...A...... ....TC....
M11             CATG.GG..A C.....A... .T.....C.. ...T...... ....AT....
MX1             CGTA.GG..A C.....A... .T.....C.. ...T...... ....AC....
                                                      STOP 4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4110       4120       4130       4140       4150
Consensus       RGAGGCACGA AGGYTVYRYC YYWAMAAYRR GGYRKMRCCT GGGAGGGSKS
TW18            A.-....... ...T.ACAT. TCT.C.-CGA ..TGTAA... .......CGC
HL4-9           A.-....... ...T.ACAT. TCT.C.-CGA ..TATAA... .......CGC
BR12            A.-....... ...T.GCGT. TCT.C.-CGA ..CGTAA... .......CGC
BZ1             A.-....... ...T.ATGC. TCT.C.-CGA ..CATAA... .......CGC
VK              A.-....... ...T.GCGT. TCT.C.-CGA ..CGTAA... .......CGC
QB              A.-....... ...T.GCGT. TCT.C.-CGA ..CGTAA... .......CGC
M11             G......... ...C.CTAT. CTA.A..TGG ..TAGCG... .......GTG
MX1             G......... ...T.CTAT. CTA.A.-TAG ..TAGCA... .......GTG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4160       4170       4180       4190       4200
Consensus       YWHTAYRSMS CCTRRGTTRK SAATMMWTWA WCWMAMYTWC TCDWWMGAGW
TW18            CAA..TGGCG ...AA-..GT G...AAA.T. T.AC.AT.A. ..TTAC...T
HL4-9           CAA..TGGCG ...AA-..GT G...AAA.T. T.AC.AT.A. ..TTAC...T
BR12            TAT..TAGCG ...AG-..GT G...CAA.T. T.AC.AC.A. ..TTAC...T
BZ1             CAC..TGGCG ...AA-..GT G...AAA.T. T.AC.AT.A. ..TTTA...T
VK              TAT..TGGCG ...GG-..GT G...AAA.T. T.AC.AC.A. ..TTAC...T
QB              CAA..TGGCG ...AA-..GT G...AAA.T. T.AC.AT.A. ..TTAC...T
M11             CAT..TGCAC ...AG...AG C...ACT.A. A.TA.CC.T. ..AAAA...A
MX1             CTA..CGCAC ...AG...AG C..-ACT.A. A.TA.CC.T. ..GAAA...A
```

```
                ....|....| ....|....| ....|....| ....|
                        4210       4220       4230
Consensus       GAGWTGRRGG MTCTGCTTWG CCCTCWCTCC TCCCA
TW18            ...A-.AG.. A.......T. .....T.... .....
HL4-9           ...A-.AG.. A.......A. .....T.... .....
BR12            ...A-.AG.. A.......T. .....T.... .....
BZ1             ...A-.AG.. A.......T. .....T.... .....
VK              ...A-.AG.. A.......T. .....T.... .....
QB              ...A-.GG.. A.......-- ---------- -----
M11             ...T-.AA.. C.......T. .....A.... .....
MX1             ...T..AG.. C.......T. .....A.... .....
```

Appendix A3b
Group III QB-like

Alignment: Allolevivirus Group III QB-like.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       GGGGGACCCC CCTTAGGGGG TCWCCTCACA CAGCAGTAYT TCACTRAGTA
TW18            .......... .......... ..A....... ........C. .....A....
HL4-9           .......... .......... ..A....... ........T. .....G....
BR12            .......... .......... ..A....... ........T. .....G....
BZ1             .......... .......... ..T....... ........T. .....G....
VK              .......... .......... ..A....... ........T. .....G....
QB              ---------- ---------- ---------- ---------- -....G....
                           ORF1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       TRAGAGGACA TATGCCTAAA TTACCRCGTG GTCTGCGTTT CGGAGCCGAT
TW18            .G........ .......... .....A.... .......... ..........
HL4-9           .G........ .......... .....A.... .......... ..........
BR12            .A........ .......... .....A.... .......... ..........
BZ1             .A........ .......... .....G.... .......... ..........
VK              .G........ .......... .....A.... .......... ..........
QB              .A........ .......... .....G.... .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       AATGAAATTC TTAATGATTT TCARGAGCTC TGGTTTCCAG AYCTCTTTAT
TW18            .......... .......... ...G...... .......... .C........
HL4-9           .......... .......... ...A...... .......... .T........
BR12            .......... .......... ...G...... .......... .T........
BZ1             .......... .......... ...G...... .......... .T........
VK              .......... .......... ...G...... .......... .T........
QB              .......... .......... ...G...... .......... .C........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       CGAATCTTCC GACACGCAYC CGTGGTACAC ACTGAARGGT CGTGTGTTGA
TW18            .......... ........T. .......... ......A... ..........
HL4-9           .......... ........T. .......... ......A... ..........
BR12            .......... ........T. .......... ......A... ..........
BZ1             .......... ........C. .......... ......A... ..........
VK              .......... ........T. .......... ......G... ..........
QB              .......... ........T. .......... ......G... ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       AYGCYCAYHT DGATGAYCGY YTACCYAATG TARGYGGTCG YCARRTMMGG
TW18            .C..T..TC. T.....T..C C....T.... ..G.T..... C..GG.CA..
HL4-9           .T..C..TC. T.....T..T C....T.... ..G.C..... T..GA.CA..
BR12            .C..T..CA. G.....C..C T....T.... ..A.T..... C..AG.AC..
BZ1             .C..T..TT. A.....T..T C....T.... ..G.C..... T..AA.AA..
VK              .C..T..TA. T.....C..T T....C.... ..A.C..... C..AG.AA..
QB              .C..C..CC. T.....T..T C....T.... ..G.C..... C..GG.AA..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       CGYACWCCRC AYCGYGYYAC YGTYCCGATT GCCTCTWCAG GCCTTCGTCC
TW18            ..T..T..A. .T..T.TT.. T..T...... ......T... ..........
HL4-9           ..T..T..A. .T..T.CT.. T..T...... ......T... ..........
BR12            ..T..T..A. .T..C.CT.. T..T...... ......A... ..........
BZ1             ..T..T..G. .C..T.CT.. C..C...... ......T... ..........
VK              ..T..A..A. .T..C.TT.. T..C...... ......T... ..........
QB              ..C..T..A. .T..C.TC.. C..T...... ......T... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       RGTAACAACC GTTCAGTATG ATCCYRCAGC ACTRTCGTTC TTRTTGAACG
TW18            G......... .......... ....CG.... ...A...... ..A.......
HL4-9           G......... .......... ....CA.... ...A...... ..A.......
BR12            A......... .......... ....CG.... ...A...... ..A.......
BZ1             G......... .......... ....TA.... ...G...... ..A.......
VK              A......... .......... ....CG.... ...A...... ..G.......
QB              G......... .......... ....CG.... ...A...... ..A.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus       CYCGTGTTGA YTGGGATTTC GGTRATGGYG AYAGTGCGRA CCTTGTCATT
TW18            .T........ T......... ...A....C. .T......A. ..........
HL4-9           .T........ T......... ...A....C. .T......A. ..........
BR12            .T........ T......... ...A....C. .C......G. ..........
BZ1             .C........ T......... ...G....T. .T......G. ..........
VK              .T........ T......... ...A....C. .C......G. ..........
QB              .T........ C......... ...A....C. .T......A. ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus       AAWGACTTYB TKTTYCGYAC YTYYGCRCCT AAGGAKTTYG ATTTYTCGAA
TW18            ..T.....TT .G..T..C.. T.CC..A... .....G..C. ....C.....
HL4-9           ..T.....TT .G..T..T.. T.TC..G... .....G..C. ....C.....
BR12            ..A.....TG .T..C..T.. C.TT..A... .....T..C. ....C.....
BZ1             ..T.....TT .G..C..C.. T.TC..A... .....G..C. ....C.....
VK              ..A.....CG .G..T..C.. C.TT..A... .....T..C. ....C.....
QB              ..T.....TC .G..T..C.. C.TT..A... .....G..T. ....T.....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       YTCYTTRGYT CCTCGYTATA CYCAGGCCTT CTCCGCBTTT AATGCCAART
TW18            T..T..A.T. .....T.... .T........ ......T... ........G.
HL4-9           T..C..A.T. .....T.... .T........ ......T... ........G.
BR12            T..C..A.C. .....T.... .C........ ......C... ........A.
BZ1             C..C..G.C. .....C.... .C........ ......T... ........G.
VK              C..C..A.T. .....T.... .C........ ......T... ........A.
QB              C..C..A.T. .....T.... .T........ ......G... ........G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus       ATGGCRYTAT GATCGGYGAA GGGCTCGAGA CTATAAAATA TCTCGGGCTK
TW18            .....AC... ......C... .......... .......... .........G
HL4-9           .....AC... ......C... .......... .......... .........T
BR12            .....GT... ......T... .......... .......... .........T
BZ1             .....AC... ......C... .......... .......... .........T
VK              .....GT... ......T... .......... .......... .........T
QB              .....AC... ......C... .......... .......... .........T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       560        570        580        590        600
Consensus       TTACTGCGCA GACTGCGTGA GGGTTWCCGY GCTGTTAARC RYGGCGATTT
TW18            .......... .......... .....A...C ........A. GC........
HL4-9           .......... .......... .....A...C ........G. GT........
BR12            .......... .......... .....A...C ........G. GC........
BZ1             .......... .......... .....T...T ........G. AC........
VK              .......... .......... .....A...T ........G. GC........
QB              .......... .......... .....A...C ........G. GT........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       610        620        630        640        650
Consensus       ACGYGCTCTT CGYAGGGTTA TCCAGTCYTA CCAYAADGGT AARTGGAARC
TW18            ...T...... ..C....... .......T.. ...T..T... ..G.....G.
HL4-9           ...T...... ..T....... .......C.. ...T..G... ..G.....A.
BR12            ...T...... ..C....... .......T.. ...T..A... ..A.....A.
BZ1             ...C...... ..T....... .......T.. ...C..G... ..A.....A.
VK              ...T...... ..T....... .......T.. ...T..A... ..A.....A.
QB              ...T...... ..T....... .......C.. ...T..T... ..G.....A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       660        670        680        690        700
Consensus       CRRCTACTGC TGGTAATCTC TGGCTTGAAT TYCGTTATGG YCTTATGCCY
TW18            .GA....... .......... .......... .T........ C........T
HL4-9           .AA....... .......... .......... .C........ T........T
BR12            .GA....... .......... .......... .T........ C........T
BZ1             .GA....... .......... .......... .T........ C........C
VK              .GA....... .......... .......... .T........ C........T
QB              .GG....... .......... .......... .T........ C........T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       710        720        730        740        750
Consensus       CTCTTYTAYG ACATCARAGA YGTYATGYTA GAYTGGCARA RBCGYCAYGA
TW18            .....C..T. ......A... C..C...T.. ..T.....G. AC..T..T..
HL4-9           .....C..T. ......A... C..C...T.. ..C.....G. AG..C..T..
BR12            .....C..C. ......G... T..T...T.. ..T.....G. GC..C..T..
BZ1             .....C..T. ......G... T..C...T.. ..T.....G. AT..T..C..
VK              .....C..T. ......G... T..C...C.. ..T.....A. AC..C..T..
QB              .....T..T. ......G... T..C...T.. ..C.....G. AC..T..T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       760        770        780        790        800
Consensus       TARRATTCAA CGCCTCCTTC GRTTTTCDGT YGGTCAYGGY GAGGAYTRYG
TW18            ..AA...... .......... .G.....T.. C.....C..T .....C.GC.
HL4-9           ..GG...... .......... .G.....G.. T.....C..T .....T.AT.
BR12            ..AG...... .......... .A.....A.. T.....T..T .....C.AC.
BZ1             ..AG...... .......... .G.....T.. C.....T..C .....T.AT.
VK              ..AG...... .......... .G.....T.. T.....C..C .....C.AT.
QB              ..AG...... .......... .G.....T.. T.....C..C .....T.AC.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       810        820        830        840        850
Consensus       TTGTCRARTT YGACARYYTR TAYCCHGCYK TWKCYTACTT TAAACTRARA
TW18            .....G.A.. C....ACT.A ..T..A..CG .TG.T..... ......G.A.
HL4-9           .....A.A.. C....GCT.A ..T..A..CG .TT.T..... ......A.G.
BR12            .....A.G.. T....ACT.G ..T..C..CT .AG.C..... ......G.A.
BZ1             .....G.A.. T....ACT.A ..T..C..CG .AG.T..... ......G.A.
VK              .....A.A.. C....ACT.A ..C..C..TT .AG.C..... ......G.A.
QB              .....G.A.. C....ATC.G ..C..T..CG .TG.T..... ......G.A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       860        870        880        890        900
Consensus       GGKGAGATTA CWCTCGAACG CCGTCATCGT CAYGGYATAT CYTACGCKAA
TW18            ..T....... .A........ .......... ..T..T.... .T.....T..
HL4-9           ..G....... .A........ .......... ..T..T.... .T.....T..
BR12            ..T....... .T........ .......... ..T..C.... .T.....G..
BZ1             ..T....... .T........ .......... ..T..T.... .T.....T..
VK              ..T....... .A........ .......... ..C..T.... .C.....G..
QB              ..G....... .A........ .......... ..T..C.... .T.....T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       910        920        930        940        950
Consensus       YCGCGRRGGA TATGCTGTTT TCGAYAACGG TTCCCTTCGG CCTGTGTCCG
TW18            T....AA... .......... ....T..... .......... ..........
HL4-9           T....AG... .......... ....T..... .......... ..........
BR12            T....AA... .......... ....T..... .......... ..........
BZ1             T....AA... .......... ....T..... .......... ..........
VK              T....GG... .......... ....C..... .......... ..........
QB              C....AA... .......... ....C..... .......... ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        960        970        980        990       1000
Consensus       ATTGGAAGGA RCTTGCCRYY GCRTTCATCA ATCCKCAHGA AGTTGCHTGG
TW18            .......... G......ACC ..A....... ....G..T.. ......T...
HL4-9           .......... A......ACT ..A....... ....T..T.. ......A...
BR12            .......... A......GTC ..G....... ....T..C.. ......T...
BZ1             .......... G......ACC ..G....... ....T..A.. ......C...
VK              .......... G......GTC ..G....... ....T..C.. ......T...
QB              .......... G......ACT ..A....... ....G..T.. ......T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1010       1020       1030       1040       1050
Consensus       GAGYTAACTC CYTACAGCTT YGTTGYYGAT TGGTTYTTGA AYGTYGGYGA
TW18            ...C...... .C........ C....TT... .....C.... .T..C..T..
HL4-9           ...T...... .C........ C....TT... .....T.... .T..C..T..
BR12            ...T...... .T........ C....CC... .....T.... .C..T..C..
BZ1             ...T...... .C........ T....TT... .....T.... .C..C..T..
VK              ...T...... .T........ T....TT... .....C.... .C..C..C..
QB              ...T...... .C........ C....TT... .....C.... .T..T..T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1060       1070       1080       1090       1100
Consensus       CATACTYGCT CARCARGGTC ARCTATATCA TAATATCGAK ATYGTAGAYG
TW18            ......T... ..A..G.... .G........ .........T ..T.....C.
HL4-9           ......T... ..G..G.... .G........ .........T ..T.....C.
BR12            ......T... ..G..A.... .G........ .........G ..C.....T.
BZ1             ......C... ..A..A.... .A........ .........T ..T.....C.
VK              ......T... ..G..A.... .A........ .........T ..T.....T.
QB              ......T... ..A..A.... .G........ .........T ..T.....C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1110       1120       1130       1140       1150
Consensus       GYTTYGACAG ACGTGACATH CGRYTCAAAT CNTTYACYAT WAAAGGTGAR
TW18            .T..T..... .........C ..GC...... .C..C..T.. A........A
HL4-9           .T..T..... .........T ..AC...... .T..C..T.. A........A
BR12            .T..C..... .........A ..GT...... .A..C..T.. T........A
BZ1             .C..C..... .........A ..GT...... .G..T..C.. A........G
VK              .T..T..... .........A ..GT...... .G..T..C.. T........A
QB              .C..T..... .........C ..GC...... .T..C..C.. A........A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1160       1170       1180       1190       1200
Consensus       CGAAATGGRC RRCCTGTTAA CGTWTCTGCB RRCYTRTCTR CTRTYGATWY
TW18            ........G. GG........ ...T.....T GA.C.G...G ..G.C...TT
HL4-9           ........G. GA........ ...T.....C GA.C.G...G ..G.C...TC
BR12            ........A. AG........ ...A.....G GA.T.A...A ..A.C...AC
BZ1             ........A. AG........ ...T.....G GA.C.G...G ..G.T...AC
VK              ........A. AG........ ...A.....G GA.C.A...G ..A.C...AC
QB              ........G. GG........ ...T.....T AG.C.G...G ..G.C...TT
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1210       1220       1230       1240       1250
Consensus       ATTTTAYAGY CGACTYCATR CKASCARBMT TCCGTTCGCT ACACTHGATC
TW18            ......C..C .....C...A .G.G..GTC. .......... .....T....
HL4-9           ......C..C .....C...A .G.C..AGC. .......... .....C....
BR12            ......T..T .....T...G .T.G..ATA. .......... .....A....
BZ1             ......T..T .....C...A .G.G..GTA. .......... .....A....
VK              ......T..T .....T...G .T.G..ACA. .......... .....T....
QB              ......C..C .....C...A .G.G..ATC. .......... .....A....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1260       1270       1280       1290       1300
Consensus       TYGATACTAC CTTTAGTTCG TWTAAACACG TYCTHGATAG TRTCKYTTTA
TW18            .C........ .......... .A........ .C..T..... .A..TT....
HL4-9           .C........ .......... .A........ .T..T..... .A..TT....
BR12            .C........ .......... .T........ .C..A..... .A..TT....
BZ1             .C........ .......... .T........ .T..C..... .A..TT....
VK              .C........ .......... .T........ .T..A..... .G..GC....
QB              .T........ .......... .T........ .T..T..... .A..TT....
                                      STOP 1                    ORF2/3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1310       1320       1330       1340       1350
Consensus       TTAACCCAAC GCRTWAAGCG YTGAAATCTT TGGGTCAATT TGATCATGGC
TW18            .......... ..A.T..... T.....-... .......... ..........
HL4-9           .......... ..A.T..... T.....-... .......... ..........
BR12            .......... ..A.T..... T.....-... .......... ..........
BZ1             .......... ..A.T..... C......... .......... ..........
VK              .......... ..A.T..... T.....-... .......... ..........
QB              .......... ..G.A..... T.....-... .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1360       1370       1380       1390       1400
Consensus       AAAATTAGAG ACTGTTACTT TARGTAACAT YGGGAAAGAT GGAMARMAAA
TW18            .......... .......... ..A....... C......... ...C.AC...
HL4-9           .......... .......... ..A....... C......... ...A.AC...
BR12            .......... .......... ..A....... C......... ...A.GC...
BZ1             .......... .......... ..A....... T......... ...C.AA...
VK              .......... .......... ..A....... C......... ...A.AC...
QB              .......... .......... ..G....... C......... ...A.AC...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1410       1420       1430       1440       1450
Consensus       CTCTGRWCCT CAAYCCRCGT GGGGTAAATC CCACYAACGG YGTTGCCKCG
TW18            .....GT... ...C..G... .......... ....T..... T......G..
HL4-9           .....GT... ...T..G... .......... ....T..... T......G..
BR12            .....GT... ...C..G... .......... ....T..... T......T..
BZ1             .....AA... ...C..A... .......... ....C..... T......T..
VK              .....GT... ...T..G... .......... ....T..... T......T..
QB              .....GT... ...T..G... .......... ....T..... C......T..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1460       1470       1480       1490       1500
Consensus       CTTTCRSAAG CGGGTGCAGT TCCTGCRYTG GAGAAGCGTG TTACCRTTTC
TW18            .....AC... .......... ......AC.. .......... .....G....
HL4-9           .....AC... .......... ......GC.. .......... .....G....
BR12            .....AG... .......... ......AT.. .......... .....G....
BZ1             .....AG... .......... ......AT.. .......... .....A....
VK              .....GG... .......... ......AT.. .......... .....G....
QB              .....AC... .......... ......GC.. .......... .....G....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1510       1520       1530       1540       1550
Consensus       GGTATCHCAG CCTTCYCGYA ATCGYAAGAA CTACAAGGTY CARGTTAARA
TW18            ......C... .....T..C. ....C..... .........C ..G.....G.
HL4-9           ......A... .....T..T. ....T..... .........C ..A.....G.
BR12            ......T... .....T..T. ....T..... .........C ..G.....A.
BZ1             ......T... .....T..T. ....T..... .........T ..G.....A.
VK              ......T... .....C..C. ....T..... .........C ..G.....A.
QB              ......T... .....T..C. ....T..... .........C ..G.....G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       TCCARAACCC RACCGCTTGY HCTGCAAACG GTTCTTGTGA CCCATCCGTT
TW18            ....G..... G........C A......... .......... ..........
HL4-9           ....G..... G........T A......... .......... ..........
BR12            ....G..... G........T C......... .......... ..........
BZ1             ....A..... A........T T......... .......... ..........
VK              ....A..... G........T T......... .......... ..........
QB              ....G..... G........C A......... .......... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1610       1620       1630       1640       1650
Consensus       ACTCGCCAGG CATATGCTGA CGTGACYTTY TCGTTCACGC AGTATAGYAC
TW18            .......... .......... ......T..C .......... .......T..
HL4-9           .......... .......... ......T..T .......... .......T..
BR12            .......... .......... ......T..T .......... .......C..
BZ1             .......... .......... ......T..T .......... .......T..
VK              .......... .......... ......C..T .......... .......T..
QB              .......... .......... ......C..T .......... .......T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1660       1670       1680       1690       1700
Consensus       CGATGAGGAA CGAGCTTTTG TTCGYACAGA GCTTRYYGCT CTGCTCGCTR
TW18            .......... .......... ....T..... ....ATC... .........A
HL4-9           .......... .......... ....T..... ....GTC... .........A
BR12            .......... .......... ....T..... ....GCT... .........G
BZ1             .......... .......... ....T..... ....GCT... .........G
VK              .......... .......... ....C..... ....GCT... .........G
QB              .......... .......... ....T..... ....GCT... .........A
```

```
                                                                                                                STOP 2
                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         1710       1720       1730       1740       1750
Consensus        RTCCTCTGYT GATYGATGCT ATYGATCRRC TGAAYCCAGC RTATTGAACR
TW18             G.......C. ...C...... ..T....AG. ....T..... G........A
HL4-9            G.......T. ...T...... ..C....AG. ....C..... G........A
BR12             G.......C. ...C...... ..C....GA. ....C..... G........A
BZ1              A.......C. ...C...... ..C....AG. ....C..... G........G
VK               G.......C. ...C...... ..C....GA. ....T..... A........A
QB               G.......C. ...C...... ..T....AG. ....C..... G........A

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         1760       1770       1780       1790       1800
Consensus        YTGCTCATTG CCGGTGGTGG CTCAGGGKMW AAMCCCGATY CGGTTRWYAT
TW18             C......... .......... .......TCA ..C......T .....ATC--
HL4-9            C......... .......... .......TCT ..A......C .....ATC..
BR12             T......... .......... .......GAA ..C......C .....GAT--
BZ1              T......... .......... .......GAA ..C......C .....GAT--
VK               T......... .......... .......GAA ..C......C .....AAT--
QB               C......... .......... .......TCA ..A......C .....ATT--

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         1810       1820       1830       1840       1850
Consensus        TCCGGATCCA CCGATYGATC CGCCGCCAGG GWCAGGYAVN TATACYTGTC
TW18             -......... .....T.... .......... .A....C.AG .....C....
HL4-9            .......... .....C.... .......... .A....T.AA .....T....
BR12             -......... .....T.... .......... .A....T.GT .....T....
BZ1              -......... .....T.... .......... .T....T.GC .....C....
VK               -......... .....T.... .......... .A....T.CC .....T....
QB               -......... .....T.... .......... .A....T.AG .....C....

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         1860       1870       1880       1890       1900
Consensus        CYTTCGCAAT TTGGTCHYTR GARGAGGTTT AYGARCCTCC TACYADKRAY
TW18             .C........ ......CC.A ..G....... .C..G..... ...T.GGA.C
HL4-9            .T........ ......CC.A ..G....... .C..G..... ...T.AGA.C
BR12             .C........ ......TC.G ..A....... .T..G..... ...C.TTG.T
BZ1              .T........ ......AT.A ..G....... .T..G..... ...C.GTG.C
VK               .T........ ......TC.A ..G....... .T..A..... ...T.GTG.C
QB               .C........ ......CC.A ..G....... .C..G..... ...T.AGA.C

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         1910       1920       1930       1940       1950
Consensus        CGACCGTGGC CTATCTAYHR YGCTRTYGAA CTCYVRYCTC GYRADTTTGA
TW18             .......... .......CAA T...G.C... ...CGGC... .CA.G.....
HL4-9            .......... .......TAA C...A.C... ...CAGC... .CA.G.....
BR12             .......... .......TCG T...G.T... ...TCGT... .CA.G.....
BZ1              .......... .......TTA T...G.C... ...CAAC... .TG.T.....
VK               .......... .......TCA T...G.T... ...CCGC... .CG.T.....
QB               .......... .......TAA T...G.T... ...CAGC... .CG.A.....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1960       1970       1980       1990       2000
Consensus       TGTTGCCCTY RAWGATCTYT TGGGYAAYAC WRADTGGCGH GAYTGGGATT
TW18            .........C G.A.....T. ....C..T.. AA.T.....C ..C.......
HL4-9           .........T A.A.....T. ....C..T.. AA.A.....C ..C.......
BR12            .........T G.T.....C. ....T..C.. TA.A.....T ..C.......
BZ1             .........T G.T.....C. ....T..C.. TA.A.....A ..C.......
VK              .........C G.T.....C. ....T..C.. TG.A.....T ..C.......
QB              .........C A.A.....T. ....C..T.. AA.G.....T ..T.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2010       2020       2030       2040       2050
Consensus       CWCGGCTBAG DTATACCACG TTYCGCGGTT GCCGTRGCAA YGGTTATATT
TW18            .A.....G.. T......... ..C....... .....G.... T.........
HL4-9           .A.....T.. T......... ..T....... .....A.... C.........
BR12            .T.....T.. G......... ..C....... .....G.... T.........
BZ1             .A.....T.. T......... ..C....... .....G.... T.........
VK              .A.....C.. A......... ..C....... .....G.... C.........
QB              .T.....T.. T......... ..C....... .....G.... T.........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2060       2070       2080       2090       2100
Consensus       GACCTYGATG CGWCTTATCT YGCTACTGAY CARGCDATGC KYGATCAGAA
TW18            .....C.... ..A....... T........T ..G..T.... GT........
HL4-9           .....T.... ..T....... T........T ..G..T.... GT........
BR12            .....T.... ..A....... C........T ..A..T.... TC........
BZ1             .....T.... ..A....... T........C ..A..G.... TT........
VK              .....T.... ..A....... C........C ..A..A.... TT........
QB              .....T.... ..A....... T........T ..G..T.... GT........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2110       2120       2130       2140       2150
Consensus       GTAYGATATT CGCRMRGGCA ARARACCYGG TGCYTTYGGY AAHRTYGARC
TW18            ...C...... ...GAG.... .A.A...T.. ...T..C..T ..CA.T..G.
HL4-9           ...C...... ...GAA.... .G.A...C.. ...T..C..T ..CG.T..G.
BR12            ...C...... ...ACG.... .A.G...T.. ...C..T..C ..AA.T..A.
BZ1             ...C...... ...ACG.... .G.A...C.. ...C..C..T ..CA.C..A.
VK              ...C...... ...ACA.... .A.G...T.. ...C..C..T ..TA.T..A.
QB              ...T...... ...GAG.... .G.A...T.. ...T..C..T ..CA.T..G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2160       2170       2180       2190       2200
Consensus       RATTYATYTA TCTTAAGTCG ATAAATGCHT ATTGYTCTCT YAGCGATATT
TW18            G...C..T.. .......... ........T. ....C..... T.........
HL4-9           G...C..T.. .......... ........T. ....C..... T.........
BR12            A...T..T.. .......... ........A. ....T..... C.........
BZ1             A...C..C.. .......... ........C. ....T..... T.........
VK              A...T..T.. .......... ........A. ....T..... T.........
QB              G...C..T.. .......... ........T. ....C..... T.........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2210       2220       2230       2240       2250
Consensus       GCGGCCTATC ACGCCGAYGG YGTRATAGTT GGCTTTTGGC GCGATCCATC
TW18            .......... .......T.. T..G...... .......... ..........
HL4-9           .......... .......T.. T..G...... .......... ..........
BR12            .......... .......C.. C..A...... .......... ..........
BZ1             .......... .......T.. T..G...... .......... ..........
VK              .......... .......T.. C..A...... .......... ..........
QB              .......... .......T.. C..G...... .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2260       2270       2280       2290       2300
Consensus       YAGTGGTGGT GCCATACCGT TTGACTTCAC DVARTTTGAT AAGACTAAAT
TW18            T......... .......... .......... TA.G...... ..........
HL4-9           T......... .......... .......... TA.G...... ..........
BR12            T......... .......... .......... TC.G...... ..........
BZ1             T......... .......... .......... AG.G...... ..........
VK              T......... .......... .......... GG.A...... ..........
QB              C......... .......... .......... TA.G...... ..........
                                                      STOP 3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2310       2320       2330       2340       2350
Consensus       GTCCDATTCA AGCYGTGATA GTCGTTCCTC GTGCTTAGTA ACTAAGGATG
TW18            ....T..... ...C...... .......... .......... ..........
HL4-9           ....T..... ...C...... .......... .......... ..........
BR12            ....T..... ...C...... .......... .......... ..........
BZ1             ....A..... ...T...... .......... .......... ..........
VK              ....G..... ...C...... .......... .......... ..........
QB              ....T..... ...C...... .......... .......... ..........
                     ORF4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2360       2370       2380       2390       2400
Consensus       AAATGCATGT CTAAGACAGC ATCYTCGCRT AACTCTCTYA GCGCACAATT
TW18            .......... .......... ...T....A. ........T. ..........
HL4-9           .......... .......... ...T....A. ........T. ..........
BR12            .......... .......... ...T....A. ........T. ..........
BZ1             .......... .......... ...C....G. ........C. ..........
VK              .......... .......... ...T....A. ........T. ..........
QB              .......... .......... ...T....G. ........C. ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2410       2420       2430       2440       2450
Consensus       GCGCCGAGCY GCGAACACAA GAATTGAGGT HGAAGGTAAC CTCGCACTKT
TW18            .........T .......... .......... A......... ........T.
HL4-9           .........T .......... .......... C......... ........T.
BR12            .........T .......... .......... T......... ........T.
BZ1             .........C .......... .......... C......... ........T.
VK              .........T .......... .......... C......... ........G.
QB              .........C .......... .......... T......... ........T.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2460       2470       2480       2490       2500
Consensus       CCATYGCCAA CGAYYTAMTR TTGGCHTATG GTCARTCGCC ATTTARYTCY
TW18            ....T..... ...TT..C.A .....T.... ....A..... .....AC..T
HL4-9           ....C..... ...TC..C.G .....T.... ....A..... .....AC..C
BR12            ....T..... ...CT..A.G .....A.... ....A..... .....GT..T
BZ1             ....T..... ...TT..A.G .....T.... ....A..... .....AT..C
VK              ....T..... ...TT..A.G .....A.... ....A..... .....GT..C
QB              ....T..... ...TT..C.G .....C.... ....G..... .....AC..T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2510       2520       2530       2540       2550
Consensus       GARKCYGAGT GTATTTCAYT CRGYCCGARA TTYGACGRRA CCCCGGAYVA
TW18            ..AG.T.... ........C. .A.T....G. ..C....GG. .......TG.
HL4-9           ..AG.T.... ........C. .A.T....G. ..T....GG. .......TC.
BR12            ..GT.T.... ........C. .A.C....A. ..C....GA. .......TA.
BZ1             ..GT.C.... ........C. .G.T....A. ..C....AA. .......CA.
VK              ..GT.T.... ........C. .A.C....A. ..C....GA. .......TA.
QB              ..GG.T.... ........T. .A.C....G. ..C....GG. .......TG.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2560       2570       2580       2590       2600
Consensus       CTTTAGGATA AATTATCTTA AAGCCGAGRT CATGTCGAAG TATGACGAYT
TW18            .......... .......... ........G. .......... ........T.
HL4-9           .......... .......... ........G. .......... ........T.
BR12            .......... .......... ........A. .......... ........T.
BZ1             .......... .......... ........A. .......... ........T.
VK              .......... .......... ........A. .......... ........T.
QB              .......... .......... ........A. .......... ........C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2610       2620       2630       2640       2650
Consensus       TCAGCCTAGG TATYGATACC GAAGCTGYTG CMTGGRARAA GTTYCTRGCA
TW18            .......... ...T...... .......C.. .C...G.A.. ...C..G...
HL4-9           .......... ...T...... .......C.. .C...G.A.. ...C..G...
BR12            .......... ...T...... .......C.. .A...A.G.. ...T..A...
BZ1             .......... ...T...... .......C.. .A...G.G.. ...C..A...
VK              .......... ...C...... .......C.. .A...G.A.. ...C..A...
QB              .......... ...T...... .......T.. .C...G.G.. ...C..G...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2660       2670       2680       2690       2700
Consensus       GCAGAGGCTG AATGTGCTTT AACGAACGCT CGYCTCTATA GRCCTRACTA
TW18            .......... .......... .......... ..T....... .G...G....
HL4-9           .......... .......... .......... ..T....... .G...G....
BR12            .......... .......... .......... ..T....... .A...A....
BZ1             .......... .......... .......... ..T....... .G...A....
VK              .......... .......... .......... ..C....... .A...A....
QB              .......... .......... .......... ..T....... .G...G....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2710       2720       2730       2740       2750
Consensus       CAGTGAGGAT TTCAATTTCT CACTGGGYGA GTCRTGTMTW CACATGGCTC
TW18            .......... .......... .......C.. ...A...A.T ..........
HL4-9           .......... .......... .......C.. ...A...A.T ..........
BR12            .......... .......... .......T.. ...G...C.T ..........
BZ1             .......... .......... .......C.. ...A...A.T ..........
VK              .......... .......... .......T.. ...A...C.T ..........
QB              .......... .......... .......C.. ...A...A.A ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2760       2770       2780       2790       2800
Consensus       GYMGAAAAAT AGYYAAGYTR ATAGGAGAYG YKCCGTCCGT TGAGGGTATG
TW18            .TA....... ..CC...T.A ........C. CT........ ..........
HL4-9           .CA....... ..TC...C.A ........T. CT........ ..........
BR12            .CC....... ..TT...T.G ........T. CT........ ..........
BZ1             .TA....... ..TT...C.A ........C. CG........ ..........
VK              .CC....... ..TT...T.A ........C. CG........ ..........
QB              .TA....... ..CC...C.A ........T. TT........ ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2810       2820       2830       2840       2850
Consensus       YTGCGYCAYT GCCGDTTTTC YGGYGGTGCY ACAACAACGA ATAACCGTTC
TW18            T....T..C. ....T..... C..T.....C .......... ..........
HL4-9           T....T..C. ....A..... T..C.....C .......... ..........
BR12            T....T..T. ....A..... T..C.....C .......... ..........
BZ1             C....T..C. ....T..... C..T.....T .......... ..........
VK              T....C..T. ....G..... T..C.....C .......... ..........
QB              T....T..C. ....A..... T..C.....T .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2860       2870       2880       2890       2900
Consensus       RYAYGGYCAT CCGTCCTTCA AGTTTGCGCT TCCRCAAGCG TGTACGCCTC
TW18            GT.C..T... .......... .......... ...A...... ..........
HL4-9           GC.C..C... .......... .......... ...A...... ..........
BR12            AT.T..T... .......... .......... ...A...... ..........
BZ1             AT.C..T... .......... .......... ...A...... ..........
VK              AT.C..T... .......... .......... ...A...... ..........
QB              GT.C..T... .......... .......... ...G...... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2910       2920       2930       2940       2950
Consensus       GGGCTTTGAA RTAYGTTYTR GCTCTYAGAG CYTCWACRCA YTTCGATATC
TW18            .......... A..T...T.A .....T.... .T..T..A.. T.........
HL4-9           .......... A..T...T.G .....T.... .T..T..A.. T.........
BR12            .......... A..T...T.G .....T.... .T..A..G.. T.........
BZ1             .......... G..T...C.G .....T.... .C..T..G.. C.........
VK              .......... A..C...T.G .....C.... .C..A..A.. T.........
QB              .......... G..T...T.A .....C.... .T..T..A.. T.........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2960       2970       2980       2990       3000
Consensus       AGRRTTTCTG ATATTAGCCC TTTTAATAAA GCAGTTACTG TACCWAAGAA
TW18            ..AA...... .......... .......... .......... ....T.....
HL4-9           ..GG...... .......... .......... .......... ....T.....
BR12            ..AG...... .......... .......... .......... ....A.....
BZ1             ..AG...... .......... .......... .......... ....A.....
VK              ..AG...... .......... .......... .......... ....A.....
QB              ..AA...... .......... .......... .......... ....T.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3010       3020       3030       3040       3050
Consensus       CAGTAARACM GATCGYTGTA TYGCTATCGA RCCYGGYTGG AATATGTTTT
TW18            ......A..C .....T.... .T........ G..T..C... ..........
HL4-9           ......A..C .....T.... .T........ A..C..C... ..........
BR12            ......A..C .....C.... .T........ A..C..T... ..........
BZ1             ......A..C .....C.... .C........ A..C..T... ..........
VK              ......G..C .....C.... .T........ A..C..T... ..........
QB              ......G..A .....T.... .T........ A..T..T... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3060       3070       3080       3090       3100
Consensus       TYCAAYTRGG YATCGGTGGC ATYYTACGYG AYCGGYTGCG TTGYTGGGGT
TW18            .T...C.A.. C......... ..TC....C. .T...T.... ...C......
HL4-9           .T...C.A.. C......... ..CC....C. .T...T.... ...C......
BR12            .C...T.A.. T......... ..TT....T. .T...T.... ...T......
BZ1             .T...C.A.. T......... ..CT....C. .C...C.... ...C......
VK              .C...T.A.. T......... ..CC....C. .T...T.... ...C......
QB              .C...C.G.. T......... ..TC....C. .T...T.... ...C......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3110       3120       3130       3140       3150
Consensus       ATCGAYCTGA AYGATCAGAC GATAAAYCAG CRYCGCGCTC ACGARGGCTC
TW18            .....T.... .T........ ......T... .GC....... ....A.....
HL4-9           .....T.... .C........ ......T... .AT....... ....A.....
BR12            .....C.... .T........ ......C... .AC....... ....G.....
BZ1             .....T.... .T........ ......C... .GT....... ....A.....
VK              .....C.... .T........ ......C... .GC....... ....A.....
QB              .....T.... .T........ ......T... .GC....... ....A.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3160       3170       3180       3190       3200
Consensus       YGTTACTAAT RAYTTAGCRA CRGTTGATCT CTCAGCRGCA AGCGATTCTA
TW18            C......... A.T.....A. .G........ ......G... ..........
HL4-9           C......... A.T.....A. .G........ ......G... ..........
BR12            C......... G.T.....A. .G........ ......G... ..........
BZ1             T......... A.C.....A. .A........ ......G... ..........
VK              C......... G.T.....G. .G........ ......A... ..........
QB              C......... A.C.....A. .G........ ......G... ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3210       3220       3230       3240       3250
Consensus       TATCTCTTGC YCTCTGYGAG CTCTTRYTRC CYCCAGGBTG GTTTGARGTY
TW18            .......... T.....T... .....AC.G. .C.....C.. ......G..C
HL4-9           .......... T.....T... .....AT.G. .C.....T.. ......G..T
BR12            .......... T.....T... .....AT.A. .T.....T.. ......G..C
BZ1             .......... T.....C... .....GT.G. .T.....G.. ......A..C
VK              .......... T.....C... .....AT.G. .T.....T.. ......G..C
QB              .......... C.....T... .....AT.G. .C.....C.. ......G..T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3260       3270       3280       3290       3300
Consensus       CTTAYGGAYC TYMGATCACC WAARGGGCRR TTGCCTRACG GTAGTGTYRT
TW18            ....T...C. .CA....... T..A....GG ......G... .......TG.
HL4-9           ....T...T. .CA....... A..A....GG ......G... .......TG.
BR12            ....C...T. .CA....... T..G....AA ......G... .......TA.
BZ1             ....T...C. .TC....... T..G....GA ......A... .......CG.
VK              ....T...C. .CA....... T..G....GA ......G... .......CA.
QB              ....T...C. .CA....... T..G....GA ......G... .......TG.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3310       3320       3330       3340       3350
Consensus       TAYYTAYGAG AARATWTCYT CBATGGGTAA YGGWTACACA TTCGARCTYG
TW18            ..CC..T... ..G..T..C. .T........ C..A...... .....G..T.
HL4-9           ..CC..T... ..A..A..C. .G........ C..A...... .....G..T.
BR12            ..CC..T... ..A..T..T. .C........ T..T...... .....G..C.
BZ1             ..CT..T... ..G..T..T. .C........ C..T...... .....A..C.
VK              ..TC..C... ..G..T..T. .C........ T..T...... .....G..T.
QB              ..CC..C... ..G..T..T. .T........ C..T...... .....G..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3360       3370       3380       3390       3400
Consensus       AGTCGCTTAT YTTYGCTTCT CTYGCYCGTT CCGTTTGTGA GATACTGRAC
TW18            .......... C..T...... ..T..T.... .......... .......G..
HL4-9           .......... C..T...... ..T..T.... .......... .......G..
BR12            .......... T..T...... ..C..C.... .......... .......G..
BZ1             .......... T..T...... ..T..T.... .......... .......G..
VK              .......... C..C...... ..T..T.... .......... .......A..
QB              .......... T..T...... ..C..T.... .......... .......G..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3410       3420       3430       3440       3450
Consensus       TTAGACTCRT CTGAGGTCAC TGTYTACGGA GAYGAYATTA TYTTRCCRTC
TW18            ........A. .......... ...C...... ..C..T.... .C..A..G..
HL4-9           ........G. .......... ...T...... ..C..T.... .T..A..G..
BR12            ........A. .......... ...T...... ..T..C.... .T..G..A..
BZ1             ........A. .......... ...C...... ..C..C.... .T..G..G..
VK              ........A. .......... ...T...... ..T..C.... .T..G..A..
QB              ........G. .......... ...T...... ..C..T.... .T..A..G..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3460       3470       3480       3490       3500
Consensus       CYGTGCAGTY CCTGCYCTYC RGGAAGTYTT YAAGTATGTT GGTTTTACDA
TW18            .T.......C .....C..T. G......C.. T......... ........A.
HL4-9           .T.......C .....C..T. G......T.. T......... ........G.
BR12            .C.......C .....T..T. A......C.. T......... ........T.
BZ1             .T.......C .....T..T. A......C.. C......... ........T.
VK              .C.......T .....T..T. A......C.. T......... ........T.
QB              .T.......C .....C..C. G......T.. T......... ........G.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3510       3520       3530       3540       3550
Consensus       CCAATACTAA RAAGACYTTY TCYGARGGGC CGTTCAGAGA GTCGTGCGGY
TW18            .......... A.....T..T ..C..G.... .......... .........T
HL4-9           .......... A.....T..C ..C..G.... .......... .........T
BR12            .......... A.....T..C ..T..G.... .......... .........T
BZ1             .......... A.....C..T ..C..A.... .......... .........T
VK              .......... G.....T..T ..T..G.... .......... .........T
QB              .......... A.....T..T ..C..G.... .......... .........C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3560       3570       3580       3590       3600
Consensus       AAGCACTACT ATTCTGGCGT AGATGTTACT CCCTTYTACA TACGYCRCCG
TW18            .......... .......... .......... .....C.... ....T.A...
HL4-9           .......... .......... .......... .....T.... ....T.A...
BR12            .......... .......... .......... .....T.... ....T.G...
BZ1             .......... .......... .......... .....T.... ....C.G...
VK              .......... .......... .......... .....T.... ....T.G...
QB              .......... .......... .......... .....T.... ....T.A...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3610       3620       3630       3640       3650
Consensus       YATAGTGAVY CCTRCCGATT TAATAYTGGT TTTGAATAAC CTATATCGGT
TW18            T.......CC ...G...... .....C.... .......... ..........
HL4-9           T.......AT ...A...... .....T.... .......... ..........
BR12            C.......GT ...G...... .....T.... .......... ..........
BZ1             C.......GT ...G...... .....T.... .......... ..........
VK              T.......GT ...G...... .....T.... .......... ..........
QB              T.......GT ...G...... .....C.... .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3660       3670       3680       3690       3700
Consensus       GGGCCACRAT TGAYGGCGTA TGGGATCCTA GRGCMCATYC YGTRTAYCTC
TW18            .......G.. ...T...... .......... .G..C...C. C..G..C...
HL4-9           .......G.. ...T...... .......... .G..C...T. C..A..C...
BR12            .......G.. ...C...... .......... .A..C...T. T..A..T...
BZ1             .......G.. ...C...... .......... .G..C...T. C..G..C...
VK              .......G.. ...C...... .......... .A..A...T. C..A..C...
QB              .......A.. ...C...... .......... .G..C...T. T..G..C...
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3710       3720       3730       3740       3750
Consensus       AAGTATCGTA AGYTGCTGCC TAAACAGCTG CARCGTAATA CTATACCTGA
TW18            .......... ..T....... .......... ..A....... ..........
HL4-9           .......... ..T....... .......... ..A....... ..........
BR12            .......... ..C....... .......... ..G....... ..........
BZ1             .......... ..T....... .......... ..G....... ..........
VK              .......... ..C....... .......... ..G....... ..........
QB              .......... ..T....... .......... ..A....... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3760       3770       3780       3790       3800
Consensus       TGGTTACGGT GATGGNGCYC TCGTCGGATC GGTCYTRATC AATCCHTTCG
TW18            .......... .....G..C. .......... ....C.G... .....T....
HL4-9           .......... .....A..C. .......... ....T.G... .....T....
BR12            .......... .....T..T. .......... ....T.G... .....T....
BZ1             .......... .....C..C. .......... ....C.G... .....A....
VK              .......... .....C..T. .......... ....T.G... .....C....
QB              .......... .....T..C. .......... ....C.A... .....T....

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3810       3820       3830       3840       3850
Consensus       CGAAAAACCG CGGRTGGATC CGGTAYGTWC CGGTGATYAY DGACCAYACA
TW18            .......... ...G...... .....C..A. .......C.C G.....T...
HL4-9           .......... ...G...... .....C..A. .......C.T A.....C...
BR12            .......... ...G...... .....T..T. .......C.T A.....T...
BZ1             .......... ...G...... .....T..A. .......C.T T.....C...
VK              .......... ...A...... .....T..T. .......C.T A.....T...
QB              .......... ...G...... .....C..A. .......T.C G.....T...

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3860       3870       3880       3890       3900
Consensus       AGGGACCGAG ARCGCRYYGA RYYRGGRTCR TATCTYTAYG AYCTYTTYTC
TW18            .......... .G...GCT.. GTTA..G..G .....T..C. .C..C..C..
HL4-9           .......... .G...GTT.. ATCG..G..G .....T..C. .C..C..T..
BR12            .......... .A...ACT.. GTCA..A..G .....C..C. .T..T..C..
BZ1             .......... .G...ACT.. GCCG..G..A .....C..T. .T..C..C..
VK              .......... .A...ACC.. GTCG..A..G .....C..C. .T..T..C..
QB              .......... .G...GCT.. GTTG..G..G .....C..C. .C..C..C..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3910       3920       3930       3940       3950
Consensus       KCGYTRYYTC YCGGAARGTA ACGATGGGTT GCCYYTWAGG GGTCCRTCGR
TW18            G..T.GTC.. C.....A... .......... ...TC.A... .....G...A
HL4-9           G..T.GTT.. T.....G... .......... ...TT.A... .....G...A
BR12            G..C.GCT.. T.....A... .......... ...TC.A... .....G...A
BZ1             T..C.GTT.. T.....A... .......... ...CC.A... .....G...A
VK              G..C.ATT.. T.....A... .......... ...TC.A... .....G...A
QB              G..T.GTC.. T.....A... .......... ...TC.T... .....A...G
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3960       3970       3980       3990       4000
Consensus       GTTGCGATYC TRYRDATCBR YYYGCYRTHG AYCARCTTAT HTGTARGAGY
TW18            ........T. .GTAA...TG TCT..TG.C. .T..A..... C....G...C
HL4-9           ........T. .GTAG...TA TCT..CA.T. .T..A..... C....G...T
BR12            ........C. .GTAT...CA CTT..CA.A. .T..G..... A....A...T
BZ1             ........T. .ATAT...GG TTT..CG.T. .C..A..... A....A...C
VK              ........T. .ATAT...CG CTC..TA.C. .T..G..... T....A...T
QB              ........T. .GCGG...TA TTT..CA.C. .T..G..... C....G...T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4010       4020       4030       4040       4050
Consensus       AATCCTACRA ARRTAAGYAG RTCTACYGGB AARTTYGATR TACAGTATAT
TW18            ........A. .GA....T.. A.....T..C ..G..C...A ..........
HL4-9           ........A. .GA....T.. A.....C..T ..G..C...A ..........
BR12            ........A. .AA....C.. A.....T..T ..G..T...G ..........
BZ1             ........A. .AG....C.. A.....T..G ..A..T...A ..........
VK              ........A. .AA....C.. A.....T..T ..G..C...A ..........
QB              ........G. .GA....C.. G.....C..C ..A..C...A ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4060       4070       4080       4090       4100
Consensus       CGCGTGCRGT AGCCGTGTTC TGGCACCCTA CGGGGTYTTC CAGGGCACGA
TW18            .......A.. .......... .......... ......C... ..........
HL4-9           .......G.. .......... .......... ......T... ..........
BR12            .......A.. .......... .......... ......C... ..........
BZ1             .......A.. .......... .......... ......T... ..........
VK              .......A.. .......... .......... ......T... ..........
QB              .......A.. .......... .......... ......C... ..........
                                     STOP 4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4110       4120       4130       4140       4150
Consensus       AGGTTRYRYC TCTACACGAG GYRTAACCTG GGAGGGCGCY AHTATRGCGC
TW18            .....ACAT. .......... .TG....... .........C .A...G....
HL4-9           .....ACAT. .......... .TA....... .........C .A...G....
BR12            .....GCGT. .......... .CG....... .........T .T...A....
BZ1             .....ATGC. .......... .CA....... .........C .C...G....
VK              .....GCGT. .......... .CG....... .........T .T...G....
QB              .....GCGT. .......... .CG....... .........C .A...G....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4160       4170       4180       4190       4200
Consensus       CTRRTTGTGA ATMAATTATC ACAAYTACTC TTWMGAGTGA GAGRGGGATC
TW18            ..AA...... ..A....... ....T..... ..AC...... ...A......
HL4-9           ..AA...... ..A....... ....T..... ..AC...... ...A......
BR12            ..AG...... ..C....... ....C..... ..AC...... ...A......
BZ1             ..AA...... ..A....... ....T..... ..TA...... ...A......
VK              ..GG...... ..A....... ....C..... ..AC...... ...A......
QB              ..AA...... ..A....... ....T..... ..AC...... ...G......
```

```
                ....|....| ....|....| ..
                        4210       4220
Consensus       TGCTTWGCCC TCTCTCCTCC CA
TW18            .....T.... .......... ..
HL4-9           .....A.... .......... ..
BR12            .....T.... .......... ..
BZ1             .....T.... .......... ..
VK              .....T.... .......... ..
QB              .....----- ---------- --
```

Appendix A4
Group IV

Alignment: Allolevivirus Group IV.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       GGGGGTAGGS SSSRTAAAGG GGGCCTGCCY TCAMCGCACT ACAGAGGAGA
BR1             -........C CCCG...--. .........T ...C...... ..........
BR8             -........C CCCG...--. .........T ...C...... ..........
HB-P22          -........C CCCG...--. .........T ...C...... ..........
HB-P24          -........C CCCA....-. .........T ...A...... ..........
SP              .........G GGGA...... .........C ...C...... ..........
NL95            .........C CCCA...--. .........T ...A...... ..........
FI              ---------- ---------- ---------- --------.. ..........
                   ORF1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       ATCTATGCCW RCCCTWCCRA KAGGWCTTCG CTTYGGWTCG AAHGGCGARR
BR1             .........A A....T..G. G...T..... ...C..A... ..T.....GG
BR8             .........A A....T..G. G...T..... ...C..A... ..T.....GG
HB-P22          .........A G....A..A. G...T..... ...T..T... ..T.....AA
HB-P24          .........A A....T..G. G...T..... ...C..A... ..C.....GG
SP              .........A A....T..G. G...T..... ...C..A... ..T.....AG
NL95            .........A A....T..G. G...A..... ...C..A... ..T.....AA
FI              .........T A....T..A. T...T..... ...T..A... ..A.....AA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       TYBTWAAYGA YTTCRRSGMG CTYTGGTTTC CRRAGCKCSW HDCHKYMGAH
BR1             .CT.A..T.. C...GAG.C. ..C....... .AG...G.CA TA.CGTC..T
BR8             .CT.A..C.. C...GAG.C. ..C....... .AG...G.CA TA.CGTA..T
HB-P22          .CT.A..C.. T...AAC.A. ..C....... .GG...T.GT CT.ATCC..A
HB-P24          .CG.T..C.. C...AAC.C. ..C....... .AG...G.GA AA.CTTC..C
SP              .TC.T..T.. C...GAG.C. ..C....... .GG...G.CA TA.CGTA..T
NL95            .CG.A..T.. T...AAC.C. ..C....... .AG...G.GA AG.ATTC..T
FI              .CT.A..T.. T...AGC.C. ..T....... .GA...G.GT CT.TTTC..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       YYNVRSHWHG GVHSCTRYRM VCTCWCKGGY TAYDTNAVTR RYCWGYYBKV
BR1             CTGAGCAAT. .GAC..GCAA G...A.T..T ..TA.C.C.A AC.T.CCTGG
BR8             CTAAGCAAT. .GAC..GCAA G...A.T..T ..TA.C.C.A AC.T.CCGGA
HB-P22          TTGAACTTA. .AAC..ATAA C...A.G..T ..CA.T.G.A AC.T.CCGGG
HB-P24          TCCGGGCTT. .ATC..ACGA A...A.T..T ..CG.A.G.A AT.A.CCGGG
SP              CTAAGCAAT. .GAC..GCAA G...A.T..T ..TA.C.C.A AC.T.CCTGG
NL95            CTCGAGCTC. .CTC..ACAC A...A.T..T ..CG.A.G.A AT.A.CCTGG
FI              TCTCAGCTT. .ACG..ACGA A...T.T..C ..TT.G.A.G GC.A.TTCTC
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       VKAYDVNRNN MNVDWSCSBA ATMMDVSVDY BMMWKBBNYK MRDACRCCGY
BR1             CT.CAGTAAT ATATTC.CG. ..AAGGGAGT CACTGTTGCT CGT..G...T
BR8             CT.CAGCAAC ACATTC.CG. ..AAGGGGGT CACTGTTGCT CGT..G...T
HB-P22          AT.CGAAGTG AAGGAG.GC. ..AAAGGCTC TCATGTGCTT CGA..A...T
HB-P24          AT.CGAGACA CGCATG.GT. ..CCTAGCAT GCATTGTATT CGT..A...C
SP              CT.CAGTGAC ATATTC.CT. ..AAAGGAGT CACTGCTGCT CGT..G...T
NL95            CT.TACGACG CGCATG.GT. ..CCTCGCAT GCATTGCGTT CGT..A...C
FI              GG.CTACGGC AGG------. ..CCTGCCTT GCAAGTCTCG AAG..A...C

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       AYMGDRSWAC NGTVCCNGTW AACCATYTYG GWTAYMGGCC NGTYACVACD
BR1             .CA.AAGT.. A..G..T..T ......C.T. .T..CC.... C..T..G..T
BR8             .CA.AAGT.. T..G..T..T ......C.C. .T..CC.... A..T..A..T
HB-P22          .CC.TAGT.. C..G..A..A ......C.T. .A..CA.... G..T..G..A
HB-P24          .CA.AAGT.. G..G..A..T ......C.T. .T..CA.... T..T..G..T
SP              .CA.AAGT.. A..G..C..T ......C.T. .T..CA.... A..T..G..T
NL95            .CA.AAGT.. T..A..T..T ......T.T. .T..TA.... A..T..G..T
FI              .TC.GGCA.. A..C..G..T ......C.T. .T..CC.... T..C..C..G

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       GTWGAGTAYR HTCCHVWBGG NACBTWCRTV CGCCTNGAYG GSVMYGTDMR
BR1             ..T.....CA T...CGAC.. C..C.A.G.G .....A..T. .GCAC..AAA
BR8             ..T.....CA T...CGAC.. C..T.A.G.G .....T..T. .GCAC..GAA
HB-P22          ..A.....TA A...ACTG.. A..T.T.A.C .....G..T. .GGAT..AAA
HB-P24          ..T.....CA C...TAAT.. G..T.T.A.C .....T..C. .GGAT..GAA
SP              ..T.....CA T...CGAC.. A..T.A.G.A .....C..T. .GCAC..GAA
NL95            ..T.....TG T...AAAC.. A..T.T.A.C .....A..T. .GGAT..AAA
FI              ..A.....CG T...CAAT.. T..G.T.G.G .....T..T. .CACC..TCG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus       VWTYDVMGGD GRVKYGGTYA RYGSSWCRST BRRDCTHRVS AAYTWYRTKR
BR1             GT.CGAA..G .GCTT...T. AT.GGT.AG. TGAT..CACG ..T.ACG.GA
BR8             GT.CGAA..G .GCTT...T. AT.GGT.AG. TGAT..CACG ..T.TCG.GA
HB-P22          GT.TTCC..A .GCTT...T. GC.CGT.GC. TAGA..TGAC ..C.ACG.TG
HB-P24          GT.TTCC..T .GAGC...C. GT.GGT.GC. CAAA..TAGC ..C.ATG.TG
SP              AT.TGAA..G .ACTT...T. AT.GGT.AG. TGAT..CACG ..T.TCG.GA
NL95            GT.TTCC..T .GATC...T. GT.GGT.GC. TAAG..AAAC ..C.TCG.TG
FI              CA.CAGC..G .AGTT...T. AT.GCA.AG. GAGG..TGAC ..C.ATA.TG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus       TYDVNYTVGC KKCNCAGGSY GGYTTCGATT ACCARTCGGT AATCGGWCCW
BR1             .CTCGT.A.. TG.T....GT ..C....... ....A..... ......A..T
BR8             .CTCAT.A.. TG.T....GT ..C....... ....G..... ......A..T
HB-P22          .CGGTC.G.. TT.G....CC ..T....... ....G..... ......A..A
HB-P24          .TAACT.G.. GG.C....GT ..C....... ....G..... ......T..A
SP              .CTCAT.A.. TG.T....GT ..C....... ....A..... ......A..T
```

```
NL95          .TAACT.A.. GT.A....GC ..C....... ....G..... ......A..A
FI            .TAACC.C.. TG.C....GT ..C....... ....G..... ......A..A
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        460        470        480        490        500
Consensus       AGGTTCTCDK CGCRMTTYKC YGCGTTYAGY ACCAARTATG GTRYBTTRCT
BR1             ........TG ...AC..CT. C.....C..C .....A.... ..GTC..A..
BR8             ........TG ...AC..CT. C.....C..C .....A.... ..GTC..G..
HB-P22          ........AT ...AA..CT. C.....T..T .....A.... ..ACT..G..
HB-P24          ........TG ...AA..CG. T.....C..T .....G.... ..ACG..A..
SP              ........TG ...GC..CT. C.....T..C .....A.... ..GTC..A..
NL95            ........GT ...AA..CT. C.....T..T .....G.... ..GCT..G..
FI              ........TT ...AA..TT. C.....T..T .....A.... ..ACT..A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        510        520        530        540        550
Consensus       MGGVGAAGGG AGAGARACDC TTARKTATCT TCTCCTSSTS BTTCGCAGRR
BR1             C..G...... .....G..T. ...AG..... ......GC.G C.......AA
BR8             C..A...... .....A..T. ...AG..... ......GC.G C.......AA
HB-P22          C..G...... .....A..A. ...GT..... ......GC.G T.......AG
HB-P24          A..C...... .....A..A. ...GT..... ......GC.G T.......GG
SP              C..A...... .....A..T. ...AG..... ......CG.C G.......AA
NL95            C..G...... .....A..A. ...GT..... ......GC.G T.......AA
FI              C..A...... .....A..G. ...GT..... ......GC.G T.......AA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        560        570        580        590        600
Consensus       TRCGTGAAGG GWWCCKYGCY GTWARGCRHG GHGATYTCAA RCGYCTCAGG
BR1             .G........ .TA..GC..C ..A.G..GT. .C...C.... G..C......
BR8             .G........ .TA..GC..C ..A.G..GT. .C...T.... G..T......
HB-P22          .A........ .AT..GC..C ..A.G..GT. .A...C.... G..T......
HB-P24          .G........ .AT..GC..T ..A.A..AC. .T...C.... G..C......
SP              .G........ .TA..GC..C ..A.G..GT. .C...C.... G..T......
NL95            .G........ .AT..GT..C ..A.G..GC. .T...C.... G..T......
FI              .G........ .TT..TC..T ..T.A..GA. .A...C.... A..C......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        610        620        630        640        650
Consensus       AATSTKRTHH SDASVTWCGA RCCDMBNWCC NHDNVHGGNM RDMDRSMAVS
BR1             ...G.GA.CT CT.CG.T... G..GACAT.. TCTAAA..GA GGAAGGC.AG
BR8             ...G.GA.CT CT.CG.T... G..GATCT.. CCGAGA..GA AGATGGC.AG
HB-P22          ...C.GG.TA GG.GC.A... A..GACGT.. GTGAAT..AC GTCGACA.GC
HB-P24          ...C.GA.AC GT.CG.T... G..GCGCT.. GTTCCC..AA AGCGGCA.AG
SP              ...G.GA.AT CG.CG.T... G..GAGTA.. ATAAAA..TA AACGAGC.AG
NL95            ...C.GA.AC GT.CA.T... A..ACGGT.. ATTTCT..CC GACGGCA.CG
FI              ...G.TA.CC GA.CG.T... A..TCGTT.. CAAGCT..GA AGAGGCA.CG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        660        670        680        690        700
Consensus       RDCNDMSTTY WMMVRRDCCT ATYGCGAYVR RMYYRYCRGW AACAAGGTYR
BR1             AT.TGAG..C TCACAAA... ..C....CAA GCTTAC.G.T ........TG
BR8             AT.AGAG..C TCACAGA... ..C....CAA ACTTAC.G.T ........TA
HB-P22          AG.CTCC..T TCAGGGA... ..C....CGA ACTTGC.A-- ----------
HB-P24          AG.ATCG..C TCAGAAG... ..C....CAG AACTGC.A-- ----------
SP              GG.CGAG..T TCACAGA... ..C....CAA GCTTAC.G.A ........CG
```

```
NL95          AA.GACG..C TCAGAGA... ..C....TCG ACTCGT.A-- ----------
FI            AT.GTCG..T AACAAGT... ..T....CAA GCTCGC.A-- ----------
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        710        720        730        740        750
Consensus       AAGTTARACC RAGTGRNGRY RRKTGGAADR RCAGTAGTGC GAGYRACCTG
BR1             ......A... A....AG.GT AAG.....TA G......... ...TG.....
BR8             ......A... A....AA.GT AAG.....TA G......... ...TG.....
HB-P22          ---------- -----AT.GC AAG.....AG A......... ...CG.....
HB-P24          ---------- -----AT.GC AGT.....AG A......... ...TG.....
SP              ......G... G....AA.GT AAG.....TA G......... ...TG.....
NL95            ---------- -----AC.AC GGG.....AG A......... ...TG.....
FI              ---------- -----GT.GC GAG.....GA A......... ...TA.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        760        770        780        790        800
Consensus       TGGYTMGART TCCGHTAYGG RYTKATGCCK TTRTTYTAYG ACATHMARTC
BR1             ...T.A..G. ....T..T.. GC.T.....G ..G..C..C. ....TC.G..
BR8             ...C.A..G. ....T..T.. AC.G.....T ..G..T..T. ....TC.G..
HB-P22          ...C.C..A. ....T..T.. GT.G.....G ..G..C..C. ....AC.A..
HB-P24          ...T.A..A. ....T..C.. GT.G.....G ..A..T..C. ....CC.G..
SP              ...T.A..G. ....T..T.. GC.G.....G ..A..C..C. ....AC.G..
NL95            ...T.A..G. ....A..C.. AC.G.....G ..G..T..C. ....CA.G..
FI              ...C.C..A. ....C..C.. AT.G.....G ..A..T..C. ....TC.G..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        810        820        830        840        850
Consensus       YKTVATGGAR GAYTTCATGC GYRTBCAYAA GARRATCGCR AARHTWCARC
BR1             CG.G.....A ..C....... .TG.C..T.. ..AG.....G ..AA.A..G.
BR8             CG.C.....A ..C....... .TG.C..T.. ..AG.....A ..GA.A..G.
HB-P22          TT.A.....A ..C....... .TG.T..T.. ..AA.....A ..AA.T..G.
HB-P24          CG.A.....A ..C....... .TG.C..T.. ..AG.....G ..AC.T..G.
SP              CG.C.....A ..C....... .TG.T..T.. ..AG.....A ..AA.T..G.
NL95            TG.A.....G ..C....... .TA.C..T.. ..AG.....G ..AT.A..G.
FI              TG.A.....A ..T....... .CG.G..C.. ..GA.....G ..AA.T..A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        860        870        880        890        900
Consensus       GRTTYTCDGC HGGRCAYGGB AAGCTNGWDR MGGTNWVBKV KMNDTTYTWY
BR1             .A..T..T.. A..G..C..T .....G.AGA C...TAGTTC GCGG..T.AC
BR8             .A..C..T.. C..A..C..T .....G.AGA C...TAGTTC GCGG..T.AC
HB-P22          .A..C..G.. T..G..C..T .....C.TTG A...GTCCGG GACT..C.AC
HB-P24          .A..C..T.. A..A..C..C .....A.TTG A...AAAGGG TAAG..C.TC
SP              .G..T..A.. T..A..T..T .....C.AGA C...TAGTTC GCGG..T.AC
NL95            .A..T..A.. A..A..C..T .....T.TGA C...TAAGGG TAGG..C.TC
FI              .A..C..T.. C..G..T..G .....A.AAA A...CTCGGA TATA..T.AT


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        910        920        930        940        950
Consensus       CCKRRYVYBY AYTTCVVBMT YGAGGTYACY GCRGTGTTRC AGCGGCGTCA
BR1             ..GGACGTCC .C...GCTC. T.....C..C ..A.....A. ..........
BR8             ..GGACGTCC .C...GCTC. T.....C..C ..A.....A. ..........
HB-P22          ..GGATGTCC .C...GGCC. T.....C..C ..A.....G. ..........
HB-P24          ..TGATCCTC .C...GCTC. T.....C..C ..A.....G. ..........
SP              ..GGACGTCC .T...AGCC. T.....C..T ..A.....A. ..........
```

```
NL95          ..TGACCCGC .C...GCTA. T.....C..C ..G.....A. ..........
FI            ..GAGCACTT .T...CAGC. C.....T..C ..A.....A. ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        960        970        980        990       1000
Consensus       TCGTTGGGGK GTNRTMTAYC AGGATACTGR TWCBTWBSCM MCYTTYRAYA
BR1             .........T ..CA.A..C. .........G .T.T.ATG.C A.T..TA.C.
BR8             .........T ..CG.A..C. .........G .T.T.ATG.C A.T..CA.T.
HB-P22          .........T ..CA.A..C. .........G .T.T.ATG.C A.T..CG.C.
HB-P24          .........G ..GA.A..C. .........G .T.G.TCG.C A.C..CA.T.
SP              .........T ..CA.A..C. .........G .T.T.TTG.C A.T..CA.C.
NL95            .........G ..AA.A..C. .........G .T.G.TGC.A C.T..CA.T.
FI              .........T ..TA.C..T. .........A .A.C.TTG.C A.C..TG.C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1010       1020       1030       1040       1050
Consensus       ATGGTCRKCT DRTCCCGGTR ARGGAYTGGM AGACRGCGGC KTTWGCAYTC
BR1             ......GT.. AG.......G .A...C...A ....A..... G..T...C..
BR8             ......GT.. AG.......G .A...C...A ....A..... G..T...C..
HB-P22          ......GG.. GG.......G .A...C...C ....A..... G..T...C..
HB-P24          ......GG.. AA.......G .G...T...C ....A..... T..A...C..
SP              ......GT.. AG.......A .A...C...A ....A..... G..T...C..
NL95            ......AG.. TA.......G .G...T...C ....A..... T..A...C..
FI              ......GG.. AA.......G .G...T...C ....G..... G..A...T..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1060       1070       1080       1090       1100
Consensus       CTTAAYCCBG CYGARRYTGC NTGGGARNTH ACWCCYTWSA GYTTCGTBGY
BR1             .....T..T. .C..AGT... A.....GG.A ..T..T.AC. .C.....T.T
BR8             .....T..C. .T..AGT... G.....GA.C ..T..C.AC. .C.....T.T
HB-P22          .....C..G. .T..AAC... A.....AT.A ..A..C.AC. .T.....T.C
HB-P24          .....T..C. .C..GAC... T.....GT.A ..T..T.AC. .T.....C.C
SP              .....T..C. .C..AGT... G.....AG.T ..T..C.AC. .C.....G.T
NL95            .....T..T. .C..AAC... T.....GT.A ..T..C.AC. .C.....T.C
FI              .....T..G. .T..AAC... C.....GC.T ..T..C.TG. .T.....C.C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1110       1120       1130       1140       1150
Consensus       RGATTGGTTT GTDAAYGTWG GKGATATGCT HGAGCAGACT SGGMCARCTY
BR1             G......... ..A..T..T. .G........ T.......-. G..C..G..T
BR8             G......... ..A..T..T. .T........ T.......-. G..C..G..T
HB-P22          G......... ..G..C..T. .T........ C.......-. C..A..A..C
HB-P24          G......... ..T..T..A. .T........ C.......-. C..C..G..C
SP              G......... ..A..T..T. .T........ T.......-. G..C..G..T
NL95            G......... ..T..T..T. .T........ C......... C..C..G..C
FI              A......... ..T..C..A. .T........ A.......-. C..C..G..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1160       1170       1180       1190       1200
Consensus       TATMGKTCHA YGTYGAYGTH GTYGACGGNT TYGAYMGRAA RGACRTRARR
BR1             ...A.G--C. C..C..T..C ..T.....T. .C..CC.G.. G...A.A.AA
BR8             ...C.G--C. C..C..T..C ..T.....T. .T..CC.G.. G...A.A.GG
HB-P22          ...C.G--C. C..C..C..T ..C.....C. .T..TA.A.. G...G.G.AA
HB-P24          ...C.--.C. T..C..C..T ..C.....G. .T..TA.A.. G...G.G.AG
SP              ...C.G--C. C..C..T..C ..T.....T. .C..CC.G.. A...A.A.AA
```

```
NL95          ...C.G..A. C..T..C..A ..C.....G. .C..TA.G.. G...G.G.AG
FI            ...C.T--T. C..C..T..A ..T.....A. .C..TA.G.. G...G.G.AG
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1210       1220       1230       1240       1250
Consensus       CTYMRDWSCR TRTCVGTRCG YGTBHTDRSD RVYGDYRBDR CRVMYDYWGC
BR1             ..CAAGTC.G .A..C..A.. C..CT.GACT GGC.ACAGTG .GCACGTA..
BR8             ..TAAGTC.A .A..C..A.. C..CT.GACT AGT.ACAGTG .ACACGTA..
HB-P22          ..CCGTAG.G .A..C..G.. C..GA.TGCG AAC.GTGCAA .GAACA---.
HB-P24          ..CAAGAG.G .G..G..A.. T..CA.AAGA GAC.ACGCGA .GCACA---.
SP              ..CAAATC.G .A..A..A.. C..GC.AACG AAC.ACGTTG .GCATGTT..
NL95            ..TAGAAG.G .A..A..G.. C..CA.AAGA GAC.ACGCAA .GACTA---.
FI              ..CAAGTC.G .G..C..A.. C..TA.AGC- GCC.TC--TG .GGACTCA..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1260       1270       1280       1290       1300
Consensus       DHVBTTYVNN YTVNSWVRRG CDVDDYTGYT GCATRGTTWY TAYTCGCGCG
BR1             TAAG..TACT C.ACGTCAA. .ACGTT..T. ....A...AT ..T.......
BR8             TAGT..TACT C.ACGTCAA. .ACGTT..T. ....A...AT ..T.......
HB-P22          ACAG..TGTC C.AAGACGA. .TAGTT..T. ....A...TC ..T.......
HB-P24          GTCT..CAGC T.AGCTGGA. .GGTGT..T. ....A...TC ..T.......
SP              TAGC..TCAG C.GCGACAA. .AAAAC..T. ....A...AT ..C.......
NL95            TAGC..CAAT T.ATCTGGA. .AAAGT..T. ....A...TC ..T.......
FI              TACT..CGAA C.CCGAAGG. .TGAGC..C. ....G...TC ..T.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1310       1320       1330       1340       1350
Consensus       TRCATACCGT YGCGTTYCCG CAAATYTCAC CACAAMTCGA YRCTGARRTM
BR1             .G........ T.....T... .....T.... .....C.... TA....GA.A
BR8             .G........ T.....T... .....T.... .....C.... TA....GA.A
HB-P22          .A........ T.....T... .....T.... .....C.... TA....GG.A
HB-P24          .G........ T.....T... .....C.... .....C.... TA....GA.A
SP              .G........ T.....T... .....T.... .....C.... TA....GA.C
NL95            .G........ T.....C... .....C.... .....C.... CA....AG.A
FI              .G........ C.....T... .....C.... .....A.... TG....GG.A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1360       1370       1380       1390       1400
Consensus       CGTAGYKTTA ARCACGTDAT CGAYAGYVTY GCMYTWYTAA CYCARCGYKT
BR1             .....CG... .G.....G.. ...T..TA.C ..AT.AT... .C..G..CG.
BR8             .....CG... .G.....G.. ...T..TA.C ..AT.AT... .C..G..CG.
HB-P22          .....CG... .G.....T.. ...C..TA.C ..CC.TC... .T..G..TT.
HB-P24          .....TG... .A.....A.. ...C..TA.T ..CC.TT... .T..G..TT.
SP              .....CG... .G.....A.. ...T..TA.C ..CC.AT... .C..A..CG.
NL95            .....CT... .G.....T.. ...C..TG.T ..CC.TT... .T..G..TT.
FI              .....CG... .G.....A.. ...C..CC.C ..CC.TT... .T..A..CT.
                STOP 1                                  ORF2/3
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1410       1420       1430       1440       1450
Consensus       TAARARRCGY TRAATYTTTG GGTCAATTYG ATCATGGCWA AATTRAATMM
BR1             ...G----.C .G..-C.... ........T. ........A. ....A...CA
BR8             ...G----.C .G..-C.... ........T. ........A. ....A...CA
HB-P22          ...A----.C .A...C.... ........T. ........A. ....A...CA
HB-P24          ...A.GA..C .A.--T.... ........C. ........A. ....G...AA
SP              ...G----.T .G..-C.... ........T. ........A. ....A...CA
NL95            ...A.AG..C .A...C.... ........C. ........A. ....G...AA
FI              ...G----.T .A...C.... ........C. ........T. ....G...AC
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1460       1470       1480       1490       1500
Consensus       GGTWACDCTT WCCRRDHTHG GWAAGRMNGS NRATMAGACT WTRACWCTWA
BR1             ...A..T... T..AAAA.C. .A...AAT.G GG..C..... T.A..T..T.
BR8             ...A..T... T..AAAC.T. .A...AAC.G GG..C..... T.A..T..T.
HB-P22          ...A..A... A..AAAT.A. .T...GCG.G CG..C..... T.A..T..T.
HB-P24          ...T..G... A..GGTA.T. .T...GCT.G AG..C..... T.A..T..T.
SP              ...A..T... T..AAAA.C. .A...AAT.G GG..C..... T.A..T..T.
NL95            ...T..G... A..GGTA.T. .T...GCT.G AA..C..... T.A..T..T.
FI              ...A..T... A..AAGC.T. .T...GAA.C TA..A..... A.G..A..A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1510       1520       1530       1540       1550
Consensus       CDCCDCGYGG SGTWAAYCCB ACKAACGGVG TGGCRWCRCT VTCYGAAGCT
BR1             .A..G..C.. G..A..C..G ..G.....C. ....GT.G.. G..T......
BR8             .A..G..C.. G..A..C..G ..G.....C. ....GT.G.. G..T......
HB-P22          .T..A..T.. G..A..T..C ..T.....A. ....GT.G.. G..C......
HB-P24          .A..G..T.. G..A..T..C ..T.....C. ....GT.G.. A..T......
SP              .A..G..C.. G..A..C..G ..G.....C. ....GT.G.. A..T......
NL95            .A..G..T.. G..A..C..T ..T.....C. ....GT.G.. A..T......
FI              .G..T..C.. C..T..T..T ..G.....G. ....AA.A.. C..C......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       GGHGCTGTDC CGGCNYTVGA RAARCGMGTR ACTGTGTCAG TYGCKCAGCC
BR1             ..A.....T. ....TT.G.. A..G..C..G .......... .T..G.....
BR8             ..A.....T. ....TT.G.. A..G..C..G .......... .C..G.....
HB-P22          ..T.....A. ....TC.C.. G..G..A..A .......... .T..T.....
HB-P24          ..A.....T. ....GT.G.. G..G..A..A .......... .T..T.....
SP              ..A.....T. ....AT.A.. G..G..C..A .......... .T..G.....
NL95            ..A.....T. ....TT.G.. G..G..A..A .......... .T..T.....
FI              ..C.....G. ....CT.G.. G..A..C..G .......... .C..G.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1610       1620       1630       1640       1650
Consensus       HTCYCGKAAY CGTAAGAAYT WYAARRTYCA RATTAARCTY CARAACCCGA
BR1             T..C..T..C ........C. TT..AG.T.. G.....A..C ..A.......
BR8             C..C..T..C ........C. TT..AG.T.. G.....A..C ..A.......
HB-P22          A..C..T..C ........C. AT..AG.T.. G.....G..C ..A.......
HB-P24          A..T..T..C ........C. AT..GA.C.. G.....A..C ..G.......
SP              A..T..G..C ........C. TT..AG.T.. G.....A..C ..A.......
NL95            A..T..T..C ........C. AT..AG.T.. G.....A..C ..G.......
FI              A..T..T..T ........T. TC..GG.C.. A.....G..T ..G.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1660       1670       1680       1690       1700
Consensus       CTGCATGCAC GARGGAYGCR TGYGACCCWT CTGTGACGCG WTCYGCKTTC
BR1             .......... ..A...C..A ..T.....A. .......... A..T..T...
BR8             .......... ..A...C..A ..T.....A. .......... A..T..T...
HB-P22          .......... ..A...C..A ..T.....A. .......... T..C..G...
HB-P24          .......... ..A...C..A ..C.....T. .......... T..C..G...
SP              .......... ..G...C..A ..T.....A. .......... A..T..T...
```

```
NL95          .......... ..A...C..A ..C.....T. .......... T..C.-G...
FI            .......... ..A...T..G ..T.....A. .......... A..T..T...
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1710       1720       1730       1740       1750
Consensus       KCRCGAYSTA ACGCTKTCGT TCACGTCRTA TTCWACYGAS SNDGARCGYG
BR1             G.A-..CG.. .....G.... .......G.. ...T..C..C GAA..A..T.
BR8             G.A-..TG.. .....G.... .......A.. ...T..T..C GCA..A..T.
HB-P22          G.G-..CC.. .....T.... .......G.. ...T..T..C GCT..A..T.
HB-P24          G.G-..CC.. .....G.... .......G.. ...T..T..C GTT..A..T.
SP              G.A-..CG.. .....G.... .......G.. ...T..C..C GAG..A..T.
NL95            T.G...CG.. .....G.... .......G.. ...A..C..G CGT..G..C.
FI              G.A-..CC.. .....G.... .......G.. ...T..T..C GAA..A..T.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1760       1770       1780       1790       1800
Consensus       CGCTRRTTCG YACTGARYTR GCRGCTCTVY TSVMRGAYVM HYTGATYRYC
BR1             ....AA.... C.....GT.G ..A.....GT .GCAG..CAA TC....CAT.
BR8             ....AA.... C.....GT.A ..G.....GT .GCAG..TCC CT....TGT.
HB-P22          ....GA.... C.....AT.G ..G.....GT .GCAA..TCC TC....TGT.
HB-P24          ....AG.... C.....AC.G ..A.....GT .GAAG..TGA TT....TGT.
SP              ....GA.... C.....AT.G ..A.....AC .GGCG..TCC AC....TGT.
NL95            ....AA.... C.....AT.A ..G.....GT .GAAG..TGA TC....TGT.
FI              ....GA.... T.....GT.A ..G.....CC .CGCG..TCC TT....TAC.
                                                  STOP 2
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1810       1820       1830       1840       1850
Consensus       GATGCDATYG AYAAYCTKAA YCCWGCNTAC TGAGCRGCGT TAYTGGYWRS
BR1             .....A..C. .T..T..G.. T..T..T... .....A.... ..C...TAGC
BR8             .....G..T. .C..C..G.. T..T..C... .....G.... ..C...TAGC
HB-P22          .....A..C. .T..T..G.. C..T..A... .....G.... ..T...CAGC
HB-P24          .....T..T. .C..T..T.. T..T..A... .....A.... ..C...CAGC
SP              .....T..T. .C..T..G.. C..A..C... .....G.... ..C...TAGC
NL95            .....A..C. .T..T..G.. T..T..A... .....A.... ..C...CAGC
FI              .....T..C. .T..T..G.. T..T..G... .....G.... ..T...TTAG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1860       1870       1880       1890       1900
Consensus       CTCGYCNRGY GGYGRGRWTA AHCCMYCMDW NCCWGRCGTY CCGGWTVKTC
BR1             ....T.CG.C ..T.G.GA.. .C..CT.AAT G..A.A...T ....A.CT..
BR8             ....T.CG.C ..T.G.GA.. .C..CT.AAA G..A.A...T ....A.CT..
HB-P22          ....T.TA.C ..T.G.GA.. .C..A---AA G..A.A...T ....A.CG..
HB-P24          ....C.GG.T ..C.G.AA.. .T..C---AA C..T.G...C ....A.AG..
SP              ....T.CG.C ..T.G.GA.. .T..CT.CGA T..A.A...C ....T.GT..
NL95            ....C.GG.T ..C.G.AA.. .T..C---TA C..T.G...C ....A.AG..
FI              ....T.AG.C ..T.A.GT.. .A..CC.AAT A..T.A...C ....A.GT..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1910       1920       1930       1940       1950
Consensus       CRRRCGTYAA RCCGCCHGRC GGWACDGGSH SSTWTMVSTG YCCSTTYDSY
BR1             .AGA...C.. A.....T.A. ..A..G..GC GC.A.ACG.. C..C..CTCC
BR8             .AGA...C.. A.....C.A. ..A..G..GC GC.A.ACG.. C..C..CGCT
HB-P22          .AGA...C.. A.....A.G. ..A..G..CA GC.A.CGC.. T..G..TACC
HB-P24          .GAA...C.. G.....T.G. ..T..A..CA CC.A.CGG.. C..G..CGCC
SP              .AGA...C.. A.....A.A. ..T..G..GC GC.A.AAG.. C..C..CGCC
NL95            .GAA...C.. G.....T.G. ..T..T..CA CC.A.CGG.. C..G..CGCC
FI              .GAG...T.. A.....T.G. ..A..G..CT CG.T.ACG.. C..G..CAGC
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1960       1970       1980       1990       2000
Consensus       TGYTAYCGCC KYGRTRVWHT HWWCRMGGHV GSBMARRASG GWKYYYSYGM
BR1             ..T..C.... TC.G.AGTA. CTA.GA..AG .GTA.GG.C. .TTCTCCT.A
BR8             ..T..C.... TC.G.AGTA. TTA.GA..TG .GTA.GG.C. .TTCTCCT.A
HB-P22          ..T..C.... TC.G.AATA. CAT.GA..TA .GGC.AA.C. .TTCTCCT.A
HB-P24          ..C..C.... GT.G.GAAC. CAT.AC..AG .CTA.GG.C. .AGCTTGT.C
SP              ..T..C.... TC.G.AGTA. TTA.GA..TC .GTA.GG.G. .TTCTCCT.A
NL95            ..C..C.... GT.G.GAAT. AAT.AC..AG .CTA.GG.C. .AGCTTGC.C
FI              ..T..T.... TC.A.ACTA. TAT.GA..CG .GCA.GG.C. .TGTCCCT.A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2010       2020       2030       2040       2050
Consensus       BMTHTAYGMV HDKGGMVVBG ARGYCHHVGT YRHKTTYGAK TAYGCKSTYG
BR1             CA.C..C.AA AGG..AGAC. .A.T.TCG.. TACT..C..T ..T..TC.C.
BR8             CA.T..C.AA AGG..AGAC. .A.T.TCG.. TACT..C..T ..C..TC.C.
HB-P22          CA.C..T.CC AGG..AGAC. .G.T.CAG.. TATG..C..T ..T..TC.C.
HB-P24          GC.T..T.CG CTT..CAGT. .G.C.ATA.. TGAG..C..T ..C..TC.T.
SP              CA.T..T.AA AGG..AGAC. .A.T.TCA.. CACT..C..T ..C..TC.C.
NL95            GC.A..C.CG TGT..CAGT. .G.C.CTA.. TGAG..C..G ..C..GC.C.
FI              TC.C..T.AA CAG..ACCG. .G.T.ACC.. TACG..T..T ..T..TG.C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2060       2070       2080       2090       2100
Consensus       AGGATTTYCT YGGKAACRHN WWYTGGCGHA ACTGGGAYVR BCGMYTRTCR
BR1             .......T.. T..G...ACA AAT.....T. .......TCA G..AT.G..A
BR8             .......C.. T..G...ACG AAT.....T. .......TCA G..AT.G..A
HB-P22          .......C.. T..G...ACC AAT.....T. .......TCA G..AC.A..A
HB-P24          .......C.. C..G...GTG TTC.....C. .......CGG T..CT.A..G
SP              .......C.. T..G...ACG AAT.....T. .......TCA G..AT.A..A
NL95            .......C.. C..G...GAG TTC.....A. .......TGG T..AT.A..A
FI              .......C.. C..T...ACT AAT.....T. .......TAG C..AC.A..A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2110       2120       2130       2140       2150
Consensus       DMWTAYGAYW TMGVBAMYCD YCGTCGTTGY MGHGGYAAYG GKTAYRTHGA
BR1             TCA..T..CT .A.CT.AT.G T........C C.T..C..C. .G..CA.T..
BR8             TCA..T..CT .A.CT.AT.G T........C C.T..C..C. .G..CA.T..
HB-P22          AAT..T..TA .A.CT.AT.G T........C C.A..C..T. .T..CG.T..
HB-P24          ACA..C..TA .C.AC.CT.A C........T C.C..T..C. .G..CG.T..
SP              GAT..T..TA .A.CT.AT.G T........C C.T..C..T. .G..CA.C..
NL95            AAA..T..TA .C.AG.CT.A T........T C.C..T..C. .G..CG.T..
FI              AAT..T..TA .C.GC.AC.T C........C A.A..T..C. .G..TG.A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2160       2170       2180       2190       2200
Consensus       YYTDRAYGCH WCHRYVATGC ARWCHGAYRV NTDYGTDYTR TCNGGYVVBT
BR1             CT.AG.T..C A.TGTG.... .GT.A..CGA A.TC..TC.G ..T..CCGT.
BR8             CT.AA.C..T A.CGCG.... .GT.A..CGA A.GT..TC.G ..T..CCGC.
HB-P22          CC.GG.C..A A.TGCG.... .GA.C..CAG C.TT..GC.A ..C..TAAG.
HB-P24          CC.AG.C..A A.CATG.... .GT.A..CGC T.AT..AT.G ..T..CGCT.
SP              CC.AG.T..A A.CGCC.... .GT.T..TGA T.TC..AT.G ..A..CCGC.
```

```
NL95          CC.TG.C..C T.CGTA.... .AT.A..CGA G.AT..AT.G ..T..TGCT.
FI            TC.TG.C..T A.AGCG.... .GT.C..CAG C.AT..GC.G ..G..CAAG.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2210       2220       2230       2240       2250
Consensus       AYSVNGTDSK NAAGRKVMWR YYNCCMRGYR YHTTYGVMKC MMYYARRTAY
BR1             .TCCA..GCG C...GTCAAG TTT..CG.CG CT..C.GCT. CATC.AG..T
BR8             .CCCG..ACG T...GTCAAG TTT..CG.CG CA..T.GCT. CATC.AG..T
HB-P22          .TCCA..GCG G...GTCAAG TTT..AG.TG CC..T.GCG. CCTC.AG..T
HB-P24          .TGAT..TGT C...ATGCAA CCA..CA.TA TC..C.ACT. CCCT.GG..C
SP              .CGGC..GCG A...GTCAAG TTT..CG.CG CC..C.GCT. AATC.AG..T
NL95            .CGAT..TGT T...ATGCAA CCG..CG.TA CC..C.ACT. CCCT.GG..C
FI              .TCGA..ACG C...GGACTA CCC..CG.TA TC..T.CAT. ACCC.GA..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2260       2270       2280       2290       2300
Consensus       YWYYTBVAHM TWMWRGRYGR YRYCTRKKTD GAYYTDKCCG MGGTRACRGC
BR1             CTCT.GA.CA .TCAA.GC.A TGC..GGT.G ..CT.AT... A...G..A..
BR8             CTCT.GA.CA .TCAA.GT.A TGC..GGT.G ..CT.AT... A...G..A..
HB-P22          CTCC.TA.CA .AAAG.AT.A TGC..GGG.G ..CT.GT... A...A..G..
HB-P24          TACC.CC.TC .AATG.AT.G TAT..ATG.G ..CT.AG... A...A..G..
SP              CTCT.GA.CA .TCAA.GT.A TGC..GGT.A ..CT.AT... A...A..A..
NL95            TACC.CC.TC .TATG.AT.G TAT..ATG.G ..CT.AG... A...A..G..
FI              TATC.TG.AC .ACAA.AT.G CGC..GGG.T ..TC.TG... C...A..G..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2310       2320       2330       2340       2350
Consensus       VTACCRYTCY TAYGGAATGG TYATYGGYTT CTGGACRGAC TCTAARAGYC
BR1             C....GT..C ..C..... .T..T..T.. ......A... .....G..C.
BR8             A....GC..C ..C..... .T..T..T.. ......A... .....A..T.
HB-P22          C....GT..C ..T..... .C..C..T.. ......G... .....G..C.
HB-P24          A....AT..T ..C..... .C..T..C.. ......A... .....G..C.
SP              G....GT..C ..C..... .T..T..T.. ......A... .....G..C.
NL95            A....GT..C ..C..... .C..T..C.. ......A... .....G..C.
FI              G....GT..C ..T..... .C..C..C.. ......A... .....G..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2360       2370       2380       2390       2400
Consensus       CGCAGMTRCC RAVYGAYTTY ACSCRDTTTR ACVGKVVKAA HTGYCCBGTW
BR1             .....C.A.. A.CC..T..C ..G.AG...A ..A.TGCT.. T..C..T..A
BR8             .....C.A.. A.CC..T..C ..G.AG...A ..A.TGCT.. C..C..T..A
HB-P22          .....C.G.. G.GT..C..C ..C.AA...G ..A.TACG.. T..T..T..T
HB-P24          .....C.G.. A.CC..C..T ..G.GT...A ..G.GCGT.. C..C..C..A
SP              .....C.A.. A.CC..T..C ..G.AG...A ..A.TGCG.. T..C..T..A
NL95            .....C.G.. A.CC..C..T ..G.GT...A ..C.GCAT.. C..C..G..A
FI              .....A.A.. G.AT..C..T ..C.GT...G ..A.TACG.. A..T..T..T
                                                 STOP 3                ORF4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2410       2420       2430       2440       2450
Consensus       CAGACSGTDA TMRTCATMCC CTCACYTYRA KYAAMTYAAR RGAGDWWGCA
BR1             .....G..G. .AA....A.. .....T.TA. GC..C.T..A G...ATA...
BR8             .....G..G. .AA....A.. .....T.TA. GC..C.T..A G...ATA...
HB-P22          .....G..G. .CG....A.. .....T.TA. GC..C.T..A G...ATA...
HB-P24          .....C..A. .CG....A.. .....T.TA. GC..C.T..A G...ATA...
SP              .....G..G. .AA....A.. .....T.TA. GC..C.T..A G...ATA...
NL95            .....C..A. .CG....A.. .....T.TA. GC..C.T..G A...GAA...
```

FI .....G..T. .CG....C.. ...-.C.CG. TT..A.C..A G...TTT...

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2460       2470       2480       2490       2500
Consensus       TGYCWAAGAC AGCWWGTCRC ARRARRRAAA TTACYCAVVY ATTGRGTAAG
BR1             ..T.A..... ...TA...G. .GA.GAG... ....T..GCT ....G.....
BR8             ..T.A..... ...TA...G. .GA.GAG... ....T..GCT ....G.....
HB-P22          ..T.T..... ...TA...A. .GA.AGA... ....T..CCT ....A.....
HB-P24          ..T.T..... ...TA...A. .AG.AGA... ....T..AGC ....A.....
SP              ..C.A..... ...TA...G. .GA.GAG... ....T..GCT ....G.....
NL95            ..T.A..... ...TT...A. .AG.AGA... ....T..AAC ....A.....
FI              ..T.T..... ...AA...G. .GA.GAG... ....C..CCT ....G.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2510       2520       2530       2540       2550
Consensus       GTCGACATMD MYTTYGAAGA CGACATCCAY WTGTCKATYG CYAAYGAYCT
BR1             ........CA AC..C..... .........C A....T..T. .C..C..C..
BR8             ........CA AC..C..... .........T A....T..T. .C..T..C..
HB-P22          ........CG AC..C..... .........T A....T..C. .C..T..T..
HB-P24          ........CA AC..T..... .........T A....T..T. .C..T..C..
SP              ........CA AC..C..... .........T A....T..T. .T..T..C..
NL95            ........CA AC..C..... .........C A....T..C. .C..C..C..
FI              ........AT CT..C..... .........T T....G..C. .C..C..C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2560       2570       2580       2590       2600
Consensus       YTKKVRGGCS TDYGGCRTHS SNVHDCTTVV BDVNGCKGAR SARTGYATYA
BR1             C.TTGA...C .AC...A.CC CGAAA...CA GAAT..T..G G.G..C..T.
BR8             C.GTGA...C .AC...A.CC CGATG...CG TGAC..T..G G.G..C..T.
HB-P22          T.TTAA...C .TT...G.TG CACCA...AC GTCA..G..G C.G..T..C.
HB-P24          C.TTGA...C .AC...A.TG CACCT...GC GTCG..G..G C.G..T..C.
SP              C.TTGA...C .AC...A.CC CTAAA...GA TTCG..G..G G.G..C..T.
NL95            C.TGCA...C .GC...G.TG CGCCA...GC GTCG..G..G C.G..T..C.
FI              T.TTAG...G .AT...G.AG GCGAA...AG CAGC..T..A G.A..C..T.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2610       2620       2630       2640       2650
Consensus       RYACYSCRTT CCCDVRYMYK KMDMWRRVYS YDGAYRMRTT YCGBRYYSMD
BR1             AC..CG.A.. ...AAGCCTT GATCAAGGCG CA..CACA.. T..TGTCGAG
BR8             AC..CG.A.. ...AAGCCTT GATCAGAGCG CA..CACG.. T..TGTCGAG
HB-P22          AC..TC.G.. ...TGACACT TCGATGACTG CT..TGCA.. T..CATCGCG
HB-P24          GC..TC.G.. ...TGACACT TCAATGGACG CT..TGCA.. T..CATCCAT
SP              AC..CG.A.. ...GAGCCTG GATCAAGGCG TT..CACG.. C..TGTCGAA
NL95            AC..TC.G.. ...TGACACT TCAATGAACC CA..TGAA.. T..GATTCAA
FI              AT..CG.A.. ...GCGTCTG GATCAGAGTC CG..TACG.. T..TACCGAG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2660       2670       2680       2690       2700
Consensus       TAYYTVMGNK CHGARATMYT NWSNAAGTWY RRYGSSCAYC CTCTHGGTAY
BR1             ..TT.GC.CG .T..G..CT. ATCG....TC GAC.GG..C. ....A....C
BR8             ..CT.AC.CG .C..G..CC. ATCG....TT GAC.GG..T. ....T....T
HB-P22          ..TC.GC.GT .A..G..CC. TAGT....AT AGT.CG..C. ....C....T
HB-P24          ..CC.GC.TT .A..G..CC. TAGC....TC AGT.CG..C. ....C....T
SP              ..CT.AC.CG .C..A..CT. ATCA....TT GAT.GG..C. ....C....T
```

```
NL95          ..TC.GC.TT .A..G..AC. CAGC....TT AGC.CG..C. ....C....T
FI            ..CC.CA.AT .C..G..CC. GTCT....TC AAT.CC..C. ....C....T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2710       2720       2730       2740       2750
Consensus       WRATACCGAA RMRRYHGCHT GGGARAARTT CCTRGCKGCY GARGAGGGHT
BR1             TG........ AAGACA..A. ....G..A.. ...G..T..T ..A.....T.
BR8             TA........ GAAGCA..A. ....G..A.. ...G..T..C ..G.....T.
HB-P22          TG........ GCAGTA..T. ....G..G.. ...A..T..T ..A.....C.
HB-P24          TG........ GCGGCC..C. ....A..G.. ...G..T..C ..G.....T.
SP              TG........ GCGGCT..A. ....A..G.. ...A..G..C ..G.....T.
NL95            TG........ GCAGCT..C. ....A..G.. ...A..G..T ..A.....A.
FI              AG........ GCAGTC..C. ....A..G.. ...A..G..T ..A.....T.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2760       2770       2780       2790       2800
Consensus       GTMGACDHAC VAAYGMRCGW CTSWCDMDVK BYAAGTACCA YGAYAATTCY
BR1             ..A...TT.. G..C.AA..A ..GT.GCAGG TT........ C..T.....C
BR8             ..A...GT.. G..C.AA..A ..GT.ACAAG CT........ T..T.....T
HB-P22          ..A...AA.. G..C.AA..A ..GT.GCAGG CT........ T..C.....C
HB-P24          ..A...AA.. C..C.AA..T ..CA.TAAGG TT........ T..T.....C
SP              ..A...AA.. G..C.AA..A ..GT.GCTAG TT........ C..T.....C
NL95            ..A...TT.. C..T.AG..T ..CA.TAAAG TT........ C..T.....C
FI              ..C...TC.. A..C.CA..A ..GT.GAGCT GC........ C..C.....C

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2810       2820       2830       2840       2850
Consensus       ATTTTRTCRT GGGGYGAGCG YGTWATYCAY ACBGCYCGYC GAAAAATACT
BR1             .....G..G. ....C..... T..A..T..T ..T..T..C. ..........
BR8             .....A..G. ....T..... C..A..T..T ..C..T..C. ..........
HB-P22          .....G..A. ....C..... T..T..C..C ..T..T..T. ..........
HB-P24          .....G..A. ....C..... T..T..T..C ..T..T..T. ..........
SP              .....G..G. ....C..... T..T..T..C ..G..C..T. ..........
NL95            .....G..G. ....C..... T..T..T..C ..T..C..T. ..........
FI              .....G..A. ....C..... T..T..T..C ..T..C..T. ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2860       2870       2880       2890       2900
Consensus       TAARCTDATH GGYGAGDCYG YNCCRYTBGG GGATGTGGCG YTRCGYKSCC
BR1             ...A..A..T ..C...T.T. TG..GT.C.. .......... T.G..CTG..
BR8             ...G..A..T ..C...T.C. TG..GT.C.. .......... T.G..TTG..
HB-P22          ...G..T..A ..C...T.T. TT..GC.C.. .......... C.G..CTG..
HB-P24          ...A..G..C ..T...A.T. TC..GC.G.. .......... T.G..CGC..
SP              ...A..A..T ..C...T.T. TA..GT.C.. .......... T.G..CTG..
NL95            ...A..G..A ..T...G.T. CG..AT.G.. .......... T.G..CGC..
FI              ...G..T..C ..C...T.T. TT..GT.T.. .......... T.A..CTG..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2910       2920       2930       2940       2950
Consensus       GDTTYTCBGG HGGYGCDACG ACCTCGGTDA ACCGTTTACA CGGYCAYCCG
BR1             .A..T..T.. C..T..A... ........A. .......... ...T..T...
BR8             .T..T..C.. C..T..A... ........G. .......... ...C..T...
HB-P22          .A..T..T.. C..C..G... ........T. .......... ...C..C...
HB-P24          .A..T..G.. A..C..T... ........T. .......... ...T..C...
SP              .T..T..T.. C..C..G... ........T. .......... ...T..T...
```

```
NL95          .A..T..G.. C..C..G... ........T. .......... ...C..C...
FI            .G..C..T.. T..C..A... ........T. .......... ...T..C...
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2960       2970       2980       2990       3000
Consensus       TCGTGGAAGC ATGCCTGTCC GCAGGATGTW ACCAAACGYG CRYTVAARTA
BR1             .......... .......... .........T ........C. .AT.C..G..
BR8             .......... .......... .........T ........C. .AC.C..G..
HB-P22          .......... .......... .........T ........C. .AT.C..A..
HB-P24          .......... .......... .........A ........T. .AC.G..A..
SP              .......... .......... .........T ........C. .AT.C..G..
NL95            .......... .......... .........A ........C. .AC.A..G..
FI              .......... .......... .........T ........C. .GC.G..G..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3010       3020       3030       3040       3050
Consensus       YCTVMWVGCB TWTAARMDRG CNTGYGGYGA YRYBGWHGAH YTMCGYRTCR
BR1             C..GCAA..C .T...ACGG. .G..C..T.. CGTT.TA..T T.A..CG..A
BR8             T..GCAA..C .T...ACGG. .A..C..T.. CGTT.TA..T T.A..CG..A
HB-P22          C..ACAA..G .A...AATG. .T..T..T.. CATC.TA..C T.A..TG..A
HB-P24          C..AATG..C .A...AAAG. .C..T..T.. CGTC.TT..T T.A..CG..A
SP              C..GCAA..C .T...GCGG. .C..T..T.. CGTT.TA..T C.A..CG..A
NL95            C..CATA..T .A...GAAA. .T..T..C.. CGTC.TA..T T.A..CG..A
FI              C..GCTC..C .A...AAAG. .G..C..C.. TACG.AC..A C.C..TA..G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3060       3070       3080       3090       3100
Consensus       RYGARGTGCG CACTTCRAAT AAAGCAGTCA CTGTTCCRAA GAAYAGYAAR
BR1             AC..G..... ......A... .......... .......G.. ...T..T..G
BR8             AC..G..... ......A... .......... .......G.. ...T..T..G
HB-P22          AT..G..... ......G... .......... .......A.. ...C..T..G
HB-P24          AT..G..... ......A... .......... .......A.. ...C..C..A
SP              AC..G..... ......A... .......... .......A.. ...C..T..A
NL95            AT..G..... ......G... .......... .......A.. ...C..T..A
FI              GC..A..... ......A... .......... .......G.. ...C..T..A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3110       3120       3130       3140       3150
Consensus       ACYGATCGYT GYATYGCYAT CGARCCYGGN TGGAAYATGT TTTTYCARYT
BR1             ..T.....C. .T..C..C.. ...A..C..T .....T.... ....C..GC.
BR8             ..T.....C. .T..C..T.. ...A..C..G .....T.... ....T..AC.
HB-P22          ..C.....T. .C..T..C.. ...A..C..A .....T.... ....C..AC.
HB-P24          ..T.....T. .C..T..C.. ...G..C..C .....T.... ....C..GT.
SP              ..T.....C. .T..T..T.. ...G..C..C .....T.... ....C..GT.
NL95            ..T.....C. .C..T..T.. ...G..C..C .....T.... ....C..GT.
FI              ..C.....C. .T..T..C.. ...G..T..C .....C.... ....C..GC.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3160       3170       3180       3190       3200
Consensus       VGGYGTMGGT GCDGTRCTMC GYGATMGGTT GCGYYTDTGG NVKATYGAYC
BR1             C..T..A... ..G..G..A. .C...A.... ...TT.A... AAG..T..C.
BR8             C..T..A... ..G..G..A. .C...A.... ...CT.A... AAG..T..T.
HB-P22          G..T..C... ..G..G..A. .C...A.... ...CC.T... GGG..T..C.
HB-P24          A..C..A... ..T..G..C. .C...A.... ...CC.T... CAG..T..T.
SP              A..C..C... ..A..G..A. .C...A.... ...TT.A... AAG..T..T.
```

```
NL95        A..C..C... ..T..G..C. .T...C.... ...CC.T... CAT..T..T.
FI          C..T..C... ..A..A..A. .T...A.... ...CT.G... TCG..C..T.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3210       3220       3230       3240       3250
Consensus       TYAATGAYCA RTCBVYYAAY CARCGYCTYG CRCGBGATGS RTCDCWGYTD
BR1             .T.....T.. G..GACC..C ..A..C..C. .G..G....G G..T.T.C.A
BR8             .T.....C.. A..GACC..T ..G..T..C. .G..G....G G..T.T.C.A
HB-P22          .T.....T.. A..CATC..T ..G..C..C. .A..C....C G..A.A.C.T
HB-P24          .T.....C.. A..TGTT..T ..G..C..C. .A..C....C G..G.A.T.G
SP              .T.....C.. A..GACC..T ..A..C..C. .G..T....G G..T.T.C.A
NL95            .C.....T.. A..TGTT..T ..G..C..C. .A..T....C A..G.A.T.G
FI              .C.....T.. A..GCTC..C ..G..C..T. .A..C....C G..A.A.T.A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3260       3270       3280       3290       3300
Consensus       RAYCATTTRG CYACYRTHGA YTTRTCNGCA GCMAGYGAYT CDATMAGYHT
BR1             A.T.....A. .T..CA.A.. C..A..C... ..C..T..T. .G..C..CT.
BR8             A.T.....A. .T..CA.A.. C..A..C... ..C..T..T. .G..A..CT.
HB-P22          G.C.....G. .C..TG.T.. C..A..C... ..A..T..T. .A..C..CC.
HB-P24          G.C.....G. .C..TG.C.. C..A..A... ..A..T..T. .G..A..CT.
SP              A.T.....A. .T..CA.A.. C..A..T... ..C..C..T. .A..C..CC.
NL95            G.C.....G. .C..TG.C.. T..A..A... ..A..C..T. .G..A..CT.
FI              G.C.....G. .C..TG.T.. C..G..G... ..C..C..C. .T..A..TA.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3310       3320       3330       3340       3350
Consensus       WMRRCTDGTB GARYTRCTVM TNCCBCCTGM NTGGTWYGRH STYYTRACVG
BR1             AAAG..A..T ..AT.G..GC .C..T....A A....AT.AC C.TC.G..A.
BR8             AAAG..A..T ..AT.G..GC .C..T....A A....AC.AC C.TT.A..G.
HB-P22          TAAG..G..G ..GT.A..AC .A..G....A C....TT.GC G.CC.G..C.
HB-P24          ACGG..T..T ..AT.G..AA .G..G....C T....TT.AT C.CC.G..C.
SP              TAAG..T..T ..GT.G..CA .G..C....A A....AT.AC C.TC.A..G.
NL95            ACGG..T..T ..AC.G..AA .G..G....C T....TT.AT C.CC.G..C.
FI              AAAA..T..C ..GT.A..GC .T..T....C G....TT.AA C.TC.A..C.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3360       3370       3380       3390       3400
Consensus       ATCTCCGATC BGAYSARGGH RTHYTRCCHR AYGGDSRWGY YGTBACYTAY
BR1             .......... T..TG.A..C G.TT.A..TG .T..ACGA.T T..G..C..T
BR8             .......... C..TC.A..C G.TT.A..TG .T..ACGA.T T..G..C..T
HB-P22          .......... G..TC.G..T A.CC.A..TG .T..GCGT.C C..T..C..C
HB-P24          .......... G..CC.G..T G.CC.G..AG .T..GCGT.T C..T..T..C
SP              .......... C..TG.A..A A.AC.G..TG .C..GCGA.T T..G..C..T
NL95            .......... G..CC.G..A A.CC.G..TG .C..GCGT.T C..T..T..C
FI              .......... G..TC.A..A G.TC.A..CA .T..TGAA.T T..C..C..C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3410       3420       3430       3440       3450
Consensus       GAGAAAATAT CCTCCATGGG WAATGGCTAC ACTTTYGARY TNGAGTCGYT
BR1             .......... .......... A......... .....C..GC .T......C.
BR8             .......... .......... A......... .....C..GC .T......C.
HB-P22          .......... .......... T......... .....C..GC .T......C.
HB-P24          .......... .......... T......... .....C..GC .G......C.
SP              .......... .......... T......... .....C..AC .C......C.
```

```
NL95          .......... .......... T......... .....T..GC .A......T.
FI            .......... .......... T......... .....C..GT .A......C.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3460       3470       3480       3490       3500
Consensus       HATHTTYGCR GCBMTMGCHM GRAGTGTGTG YGAGTTRYTR GAHMTYGACC
BR1             T..T..C..G ..GC.C..AA .A........ T.....AC.A ..AA.C....
BR8             T..T..C..G ..GC.C..AA .A........ C.....AC.A ..AA.C....
HB-P22          C..A..T..A ..GA.C..TC .G........ C.....AC.G ..TC.T....
HB-P24          A..C..T..G ..CC.A..CA .G........ C.....GT.G ..CC.T....
SP              T..T..T..G ..TA.C..TC .A........ C.....AC.G ..AA.T....
NL95            A..T..C..G ..TC.C..CA .A........ C.....AT.G ..CC.T....
FI              T..C..T..G ..TA.C..TC .G........ T.....AC.G ..CC.C....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3510       3520       3530       3540       3550
Consensus       ARTCHRCYGT YAGCGTSTAY GGDGATGATA TWATCATCSM YWCNSRTGCY
BR1             .A..TA.T.. T.....G..C ..G....... .A......GA CA.TCG...C
BR8             .A..TA.T.. T.....G..C ..G....... .A......GA CA.TCG...C
HB-P22          .A..AA.T.. C.....G..C ..A....... .A......GA TT.CCG...T
HB-P24          .G..AA.C.. C.....G..C ..T....... .A......GA CT.ACG...C
SP              .A..TA.T.. T.....G..C ..G....... .A......GA TA.CCG...C
NL95            .G..AA.T.. C.....G..C ..T....... .A......GA TT.ACG...C
FI              .G..CG.C.. C.....C..T ..A....... .T......CC CT.GGA...T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3560       3570       3580       3590       3600
Consensus       GCDSMDVYMY TDATGGMKGT BTTYGARTAY GTYGGSTTYA CBSCKAAYMR
BR1             ..AGCTCCAT .A....AT.. C..T..A..C ..T..G..C. .GC.T..CAA
BR8             ..AGCTCCAT .A....AT.. C..C..A..C ..T..G..T. .GC.T..CAA
HB-P22          ..TGCAACCC .T....AT.. G..C..G..C ..T..G..C. .CC.T..CAG
HB-P24          ..TGATGTCC .G....CG.. T..T..G..T ..T..G..C. .GC.T..CAG
SP              ..AGCTCCAT .A....AT.. C..T..G..C ..C..G..C. .TC.T..CAG
NL95            ..TGATGTCC .T....CG.. T..C..G..T ..T..G..T. .GC.T..TCG
FI              ..GCAGACCC .G....AT.. C..C..G..T ..T..C..T. .GG.G..CAG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3610       3620       3630       3640       3650
Consensus       RAAGAARACB TTYDKHRVBG GHCCSTTYMG VGARTCSTGY GGWAARCACT
BR1             A.....G..G ..CTGCGAT. .A..C..CC. C..A..G..C ..T..G....
BR8             G.....A..G ..CTGCGAT. .A..C..CC. C..A..G..C ..T..G....
HB-P22          A.....G..C ..TGTAAGT. .T..C..TA. A..G..G..C ..A..G....
HB-P24          A.....G..T ..CATTAAG. .C..C..TA. A..G..G..C ..T..G....
SP              A.....A..G ..CTGCGAT. .A..C..CC. C..A..G..C ..T..G....
NL95            A.....A..T ..CATTAAG. .C..C..TA. A..G..G..C ..A..G....
FI              A.....A..T ..CATCACC. .C..G..CC. G..A..C..T ..A..A....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3660       3670       3680       3690       3700
Consensus       GGYWYYHHGG KGTDGAYGTR ACGCCYTTTT ACATACGHCG MCCRATHCGY
BR1             ..TTCCAA.. G..A..T..A .....C.... .......C.. A..A..T..T
BR8             ..TTCCAA.. G..A..T..A .....C.... .......C.. A..A..T..T
HB-P22          ..CACTCT.. G..T..T..A .....T.... .......C.. C..A..A..T
HB-P24          ..CACTCC.. G..T..C..A .....C.... .......C.. C..A..C..T
SP              ..TTCCAA.. G..A..T..A .....C.... .......A.. A..A..A..T
```

```
NL95          ..CACTCC.. G..T..C..A .....C.... .......C.. C..A..C..C
FI            ..TTTCTC.. T..G..C..G .....T.... .......T.. C..G..A..T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3710       3720       3730       3740       3750
Consensus       WGYCTWGYYG AYATGATAYT BGTWYTRAAY ARYMTMTAYM GRTGGGGYAC
BR1             T.C..T.CT. .T......C. G..TT.A..C .GTA.C..CA .G.....C..
BR8             T.C..T.CT. .T......C. C..TT.G..T .GTA.C..CA .A.....C..
HB-P22          T.C..A.TC. .C......C. T..TC.G..T .GTA.C..TC .G.....C..
HB-P24          T.C..A.CC. .C......C. T..TT.A..C .GTA.A..TC .G.....C..
SP              T.C..A.CC. .T......C. T..AT.A..T .GTA.C..TA .G.....C..
NL95            T.C..A.CC. .C......C. T..AT.G..C .GTA.C..CC .G.....T..
FI              A.T..T.CC. .T......T. G..TT.G..T .ACC.C..TC .G.....C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3760       3770       3780       3790       3800
Consensus       KRTYGAYGGB RTRTGGGATC CTAGRGYVCT GMCYGTDTAY SARAAGTAYS
BR1             TG.T..T..T G.A....... ....A.CA.. .A.C..G..C G.A.....TC
BR8             TG.T..T..C G.G....... ....A.CA.. .A.C..A..C G.A.....TC
HB-P22          GA.T..C..C G.A....... ....G.TG.. .C.C..A..C C.G.....CG
HB-P24          GA.T..C..G G.G....... ....A.TA.. .C.T..G..C C.G.....CG
SP              TG.T..T..C A.A....... ....A.CA.. .A.C..T..C G.A.....TC
NL95            GA.T..C..T G.G....... ....G.TA.. .C.C..A..T C.G.....TG
FI              TG.C..C..T G.A....... ....A.CC.. .A.C..A..C C.G.....CG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3810       3820       3830       3840       3850
Consensus       TBAADHTSCT RCCRAGRRAT TGGMGDMGBA AYMSDATACC DGAYGGYTAC
BR1             .C..GT.G.. G..A..AA.. ...C.AC.C. .CCGG..... A..C..C...
BR8             .C..AT.G.. G..A..AA.. ...C.AC.C. .CCGG..... G..C..C...
HB-P22          .G..TA.G.. G..G..AA.. ...C.GC.T. .CACG..... T..T..T...
HB-P24          .G..GC.G.. A..A..AG.. ...C.GC.C. .CACT..... T..T..T...
SP              .T..AC.G.. G..A..AA.. ...C.TC.C. .TCGG..... A..C..C...
NL95            .G..AT.G.. A..A..AG.. ...C.GC.T. .CACA..... T..T..T...
FI              .G..GC.C.. A..A..GA.. ...A.GA.G. .CACG..... T..T..T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3860       3870       3880       3890       3900
Consensus       GGWGACGGNG CYCTCGTCGG ATYKGCYWYR ACKAAYCCRT TYGTWMTMGT
BR1             ..A.....A. .T........ ..TG..TACG ..G..C..G. .C..TA.A..
BR8             ..A.....A. .T........ ..TG..TACG ..G..C..G. .C..TA.A..
HB-P22          ..A.....C. .T........ ..TG..TACA ..G..T..G. .T..TA.A..
HB-P24          ..T.....T. .T........ ..TG..CACG ..T..C..A. .C..TA.C..
SP              ..A.....A. .T........ ..TG..TACG ..G..C..G. .T..AA.A..
NL95            ..A.....T. .C........ ..TG..CACA ..T..C..G. .T..TA.C..
FI              ..A.....G. .T........ ..CT..CTTA ..G..C..G. .C..AC.C..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3910       3920       3930       3940       3950
Consensus       BMRRAAYTWY YMVMGMBDRT AYCCKGTRTT RGTTGARGTM CAGARGGAYR
BR1             TAAG..T.AT TCAA.ACTA. .C..G..A.. A.....A..C ....G...CG
BR8             TAAG..T.AT TCAA.ACTA. .C..G..A.. A.....A..C ....G...CG
HB-P22          CAAG..T.TC TCGA.ACTA. .C..G..A.. A.....A..C ....A...CG
HB-P24          CCGG..T.AT TCCA.ATGG. .T..T..G.. G.....G..C ....G...TG
SP              TAAA..T.AT TCAA.ACTA. .C..G..A.. A.....A..C ....G...CG
```

```
NL95          CCGG..T.AT TCCA.ATGG. .T..T..G.. G.....G..C ....G...TG
FI            GCGA..C.TC CAGC.CGAG. .C..G..G.. A.....G..A ....A...TA
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3960       3970       3980       3990       4000
Consensus       YHMMRCGHMR CGAGDWNGGY RSKTAYYTRT AYBMCHTVCT NCGYRAYCGY
BR1             TCAAG..TAG ....GTA..C AGT..TC.A. .TGC.T.A.. C..CG.T..C
BR8             TCAAG..CAG ....GTA..T AGT..TC.G. .CTC.C.A.. C..CA.T..C
HB-P22          TCAAG..TCA ....TAC..C AGT..TT.A. .TGC.A.G.. A..CG.C..C
HB-P24          CTAAG..TCA ....TTT..C AGT..CT.A. .TGC.T.G.. G..TG.C..C
SP              TCAAG..CAG ....GAG..T AGT..TC.A. .TGC.C.C.. A..TG.T..C
NL95            CAAAA..ACA ....TTC..C AGT..CT.A. .TGC.T.A.. A..TG.T..T
FI              CCCCA..TAG ....AAA..T GCG..TC.A. .TCA.C.A.. T..CG.T..T

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4010       4020       4030       4040       4050
Consensus       GADRCACGTY AYARYCCBTT CCTGYRTRHY RCAGATDGGR BTRGKTTTGA
BR1             ..GA.....T .T.GT..T.. ....CG.GAC G.....A..A C.G.G.....
BR8             ..GA.....T .C.GT..T.. ....CG.GAC G.....A..A C.G.T.....
HB-P22          ..GA.....T .C.AT..C.. ....CG.GTC G.....G..A C.G.G.....
HB-P24          ..AG.....T .C.AT..C.. ....CG.ACT G.....G..A G.G.G.....
SP              ..GA.....T .C.GT..T.. ....CG.GAC G.....A..A C.G.T.....
NL95            ..TG.....T .C.GT..T.. ....CG.GTC G.....G..A C.G.G.....
FI              ..GG.....C .C.AC..G.. ....TA.GAT A.....T..G T.A.G.....

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4060       4070       4080       4090       4100
Consensus       TGAARCRCCR CTAGCTACTA GMCTTCGTCG YAAGACWGGT CGGTACAAAG
BR1             ....G.G..G .......... .C........ C.....A... ..........
BR8             ....G.G..G .......... .C........ C.....A... ..........
HB-P22          ....G.A..G .......... .C........ C.....A... ..........
HB-P24          ....A.A..G .......... .C........ C.....A... ..........
SP              ....G.G..G .......... .C........ C.....A... ..........
NL95            ....G.G..A .......... .C........ C.....A... ..........
FI              ....G.A..G .......... .A........ T.....T... ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4110       4120       4130       4140       4150
Consensus       TDGCGTGGAT TCAGGACAGT GCCTTCATCC GGCCCCCKTA TTYNMTTAMG
BR1             .G........ .......... .......... .......G.. ..TCA...C.
BR8             .A........ .......... .......... .......G.. ..TCA...C.
HB-P22          .G........ .......... .......... .......G.. ..TTC...C.
HB-P24          .G........ .......... .......... .......T.. ..TTA...A.
SP              .G........ .......... .......... .......G.. ..TAA...C.
NL95            .T........ .......... .......... .......T.. ..TTA...A.
FI              .G........ .......... .......... .......G.. ..CG---.C.
                                                  STOP 4
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4160       4170       4180       4190       4200
Consensus       GGMMTTCCCG AGGTGAAGYT YGCDMGCWAG RCACTAGYKH GTRATGGGAA
BR1             ..CA...... ........C. T..AA..T.. A......CTT ..G.......
BR8             ..CA...... ........C. T..AA..T.. A......CTT ..G.......
HB-P22          ..AC...... ........C. T..TA..T.. A......CTA ..A.......
HB-P24          ..AA...... ........C. T..GA..T.. A......CTC ..A.......
SP              ..AA...... ........C. C..AA..T.. G......CTT ..G.......
NL95            ..AA...... ........C. T..AA..T.. A......CTT ..A.......
```

FI ..AC..... ....T. T..GC..A.. A.....TGC ..A.....

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        4210       4220       4230       4240       4250
Consensus       RGGBSBYMKY TGASSBCCVR RRRRGAGAAA GAAAGRAAAC TCCCYYYHGH
BR1             G..CGGTCTT ...CCG..CG AGAG..... .....A.... ....CTCA.T
BR8             G..CGGTCTT ...CCG..CG AGAG..... .....A.... ....CTCA.T
HB-P22          G..GCTCATT ...GGC..CG GAAG..... .....A.... ....TCCT.A
HB-P24          G..GCCTCTT ...GGT..CG GAAG..... .....A.... ....TCCT.C
SP              G..TGGTCTC ...CCG..CG AGAG..... .....G.... ....CTCC.C
NL95            G..GCCTCGT ...GGC..AG GAAG..... .....A.... ....TCCT.A
FI              A..GGCTCGT ...GCC..GA GAGA..... .....A.... ....TCTT-T


                ....|....| ....|....| ....|....| ....
                        4260       4270       4280
Consensus       GAGRGKKGGC TCTGCTTTGC CCWCTCTCCT CCCA
BR1             ...G.TG... .......... ..A....... ....
BR8             ...G.TG... .......... ..A....... ....
HB-P22          ...G.TG... .......... ..T....... ....
HB-P24          ...G.TG... .......... ..A....... ....
SP              ...G.TG... .......... ..A....... ....
NL95            ...G.TG... .......... ..A....... ....
FI              ...A.GT... .......... ---------- ----
```

Alignment of Amino Acid Sequences for each Gene
Appendix B1
Group I

Amino acid sequences of *Leviviridae* Group I. Note the YGDD motif in all *Leviviridae* replicase proteins.

Group I isolates DL1, DL2, DL13, DL16, J20, ST4, R17, M12, MS2 and fr. GenBank strain M12 sequences were not complete; therefore, the replicase gene was omitted. The most conserved stretch of amino acids are highlighted in red. Gaps indicate degenerative sites.

Group I maturation protein:

Alignment: Align Group I maturation protein.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MR F           V   R  YA G  EDNS   L YRSNW  PG   STG
DL1 matura      MRAFSVLDKE SETFVPLVRT YADGEVEDNS FSLKYRSNWT PGRFNSTGAR
DL2 matura      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL13 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL16 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
J20 matura      MRAFSVLDRE SETFVPSVRV YADGEVEDNS FSLKYRSNWT PGRFNSTGTR
ST4 matura      MRAFSTLDRE NETFVPSVRV YADGETEDNS FSLKYRSNWT PGRFNSTGAK
R17 matura      MRAFSALDKE SKTFVPSIRV YANGETEDNS FSLKYRSNWT PGRFNSTGAR
M12 matura      MRAFSVLDQE NETFVPSVRV YADGETEDNS FSLKYRSNWT PGRFNSTGAR
MS2 matura      MRAFSTLDRE NETFVPSVRV YADGETEDNS FSLKYRSNWT PGRFNSTGAK
fr maturat      MRKFIPTERM SKSHVVSVRE YADGELEDNS LPLIYRSNWS PGQYTSTGPR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       T  WHYPS Y SRGA       DQG Y R G  SWGR  EE   G G S DARS
DL1 matura      TKQWHYPSSY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GYGFSLDARS
DL2 matura      TKQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL13 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL16 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
J20 matura      TNQWHYPSPY SRGALSVTSV DQGSYKRSGS SWGRPYEEKA GFGFSLDARS
ST4 matura      TKQWHYPSPY SRGALSVTSI DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
R17 matura      TKQWHYPSPY SRGALSVTSI DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
M12 matura      TKQWHYPSPY SRGALSVTAI DQGAYKRSGS SWGRPYEEKT GFGFSLDARS
MS2 matura      TKQWHYPSPY SRGALSVTSI DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
fr maturat      TKEWHYPSSY SRGAIGIKAL DQGKYARLGT SWGREFEERA GYGMSIDARS
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         110        120        130        140        150
Consensus        CYSLFPVSQN  T I VP NV ANRA TEVL  KVTQGNFNLG VALAEARSTA
DL1 matura       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL2 matura       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL13 matur       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL16 matur       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
J20 matura       CYSLFPVSQN MTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
ST4 matura       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
R17 matura       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
M12 matura       CYSLFPVSQN LTYIEVPQNV ANRATTEVLQ KVTQGNFNLG VALAEARSTA
MS2 matura       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
fr maturat       CYSLFPVSQN LTWIDVPTNV ANRATTEVLG KVTQGNFNLG VALAEARSTA


                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         160        170        180        190        200
Consensus        SQL TQTIAL  KAYTAARRG NWRQ  RYLA LNE RKF SK  VA RWLELQ
DL1 matura       SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL2 matura       SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL13 matur       SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL16 matur       SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
J20 matura       SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
ST4 matura       SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
R17 matura       SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
M12 matura       SQLATQTIAL VKAYTAARRG NWRQPVRYLA LNEDRKFRSK HVAGRWLELQ
MS2 matura       SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
fr maturat       SQLSTQTIAL IKAYTAARRG NWRQALRYLA LNENRKFNSK SVASRWLELQ


                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         210        220        230        240        250
Consensus        FGW PL SDI QGAYEMLTKV HL  F PMRA VRQVG N  L  GRL  PAA
DL1 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNVKL DGRLSYPAAN
DL2 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL13 matur       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL16 matur       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
J20 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNVKL DGRLSYPAAN
ST4 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL NGRLSYPAAN
R17 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLAYPAAN
M12 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
MS2 matura       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
fr maturat       FGWMPLLSDI QGAYEMLTKV HLKAFMPMRA VRQVGQNVSL SGRLTSPAAS


                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         260        270        280        290        300
Consensus           TCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSF VDWL
DL1 matura       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL2 matura       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL13 matur       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL16 matur       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
J20 matura       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
ST4 matura       FQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
R17 matura       FQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
M12 matura       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
```

MS2 matura FQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
fr maturat YKSTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFLVDWL

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       LPVGNMLEGL TAP GCSY S GTVTDVI GE S I  D  YG W   R  TAK
DL1 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
DL2 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL13 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL16 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
J20 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
ST4 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
R17 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
M12 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
MS2 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
fr maturat      LPVGNMLEGL TAPIGCSYQS GTVTDVISGE STITADDIYG WDTVRPATAK


                ....|....| ....|....| ....|....| ....|....| ...
                       360        370        380        390
Consensus          SA HRGV QSV PTTG Y VKSPFS VHT LDALAL RQR L
DL1 matura      AQISAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LSR
DL2 matura      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL13 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL16 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
J20 matura      AQISAMHRGV QSVWPTTGVY VKSPFSIVHT LDALALIRQR LSR
ST4 matura      AQISAMHRGV QSVWPTTGAY VKSPFSMVHT LDALALIRQR LSR
R17 matura      AHVSAMHRGV QSVWPTTGAY VKSPFSMVHT LDALALIRQR LSK
M12 matura      AQVSAMHRGV QSVCPTTGVY VKSPFSMVHT LDALALIRQR LSK
MS2 matura      AQISAMHRGV QSVWPTTGAY VKSPFSMVHT LDALALIRQR LSR
fr maturat      VQISAVHRGV QSVWPTTGVY VKSPFSMVHT LDALALFRQR LWK
```

Group I capsid protein:

Alignment: Align Group I capsid.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MASNF  FVL VDNGGTGDV  V PSNFANGV AEWISSNSRS QAYKVTCSVR
DL1 coat        MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL2 coat        MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL13 coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL16 coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
J20  coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
ST4 coat        MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
R17 coat        MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
M12 coat        MASNFTQFVL VDNGGTGDVT VXPSNFANGV AEWISSNSRS QAYKVTCSVR
MS2 coat        MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
fr coat         MASNFEEFVL VDNGGTGDVK VAPSNFANGV AEWISSNSRS QAYKVTCSVR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       QSSA NRKYT  KVEVPKVAT Q  GGV LPV AAWRSY NME LTIP FATN
DL1 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL2 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL13 coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL16 coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
J20  coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
ST4 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
R17 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
M12 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
MS2 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
fr coat         QSSANNRKYT VKVEVPKVAT QVQGGVELPV AAWRSYMNME LTIPVFATND

                ....|....| ....|....| ....|....|
                        110        120        130
Consensus       DC LIVKA Q G  K GNPI   AIAANSGIY
DL1 coat        DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL2 coat        DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL13 coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL16 coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
J20  coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
ST4 coat        DCELIVKAMQ GLLKDGNPIP SAIAANSGIY
R17 coat        DCELIVKAMQ GLLKDGNPIP SAIAANSGIY
M12 coat        DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
MS2 coat        DCELIVKAMQ GLLKDGNPIP SAIAANSGIY
fr coat         DCALIVKALQ GTFKTGNPIA TAIAANSGIY
```

Group I lysis protein:

Alignment: Align Group I lysis.

```
              ....|....| ....|....| ....|....| ....|....| ....|....|
                       10         20         30         40         50
Consensus     M       SQ  T  S      PF HE YPC   QQRSSTLYV LI LAIFLSK
DL1 lysis     METRSPQQSQ QTPESTNRFR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL2 lysis     METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL13 lysis    METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL16 lysis    METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
ST4 lysis     METRFPQQSQ QTPASTNRRR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
R17 lysis     METRFPQQSQ QTPASTNRCR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
J20 lysis     METQSPQQSQ PTPESINRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
M12 lysis     METRFLRQSQ QTPASTNRYR PFKHEDYPCR XQQRSSTLYV LIFLAIFLSK
MS2 lysis     METRFPQQSQ QTPASTNRRR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
fr lysis      MQ----QPSQ PTRESTKKPV PFQHEEYPCQ NQQRSSTLYV LICLAIFLSK


              ....|....| ....|....| ....|
                       60         70
Consensus     FTNQLL SLL    IR V T   QLLT
DL1 lysis     FTNQLLLSLL EAVIRTVETL QQLLT
DL2 lysis     FTNQLLLSLL EAVIRTVETL QQLLT
DL13 lysis    FTNQLLLSLL EAVIRTVETL QQLLT
DL16 lysis    FTNQLLLSLL EAVIRTVETL QQLLT
ST4 lysis     FTNQLLLSLL EAVIRTVTTL QQLLT
R17 lysis     FTNQLLLSLL EAVIRTVTTL QQLLT
J20 lysis     FTNQLLLSLL EAVIRTVETL RQLLT
M12 lysis     FTNQLLLSLL DAVIRTVTTF QQLLT
MS2 lysis     FTNQLLLSLL EAVIRTVTTL QQLLT
fr lysis      FTNQLLASLL DLLIRIVTTL QQLLT
```

Group I replicase protein:

Alignment: Align Group I replicase.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MSK TKKFNS LCIDL  DLS LE YQSIASV ATGS  PHS  DFTAIAYLRD
DL1 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL2 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL13 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL16 repli      MSKTTKKFNS LCIDLPCDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
ST4 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSD DFTAIAYLRD
R17 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGNPHSD DFTAIAYLRD
J20 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
MS2 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSD DFTAIAYLRD
fr replica      MSKSTKKFNS LCIDLSRDLS LEVYQSIASV ATGSSDPHSD DFTAIAYLRD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       ELLTKHP LG  GNDEATRR  LAIAKL EAN    GQINR G FLHD   SWD
DL1 replic      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL2 replic      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL13 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL16 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
ST4 replic      ELLTKHPTLG SGNDEATRRT LAIAKLREAN DRCGQINREG FLHDKSLSWD
R17 replic      ELLTKHPTLG SGNDEATRRT LAIAKLREAN DRCGQINREG FLHDKSLSWD
J20 replic      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
MS2 replic      ELLTKHPTLG SGNDEATRRT LAIAKLREAN GDRGQINREG FLHDKSLSWD
fr replica      ELLTKHPNLG DGNDEATRRS LAIAKLLEAN DRCGQINRDG FLHDATASWD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       PDVLQTSIRS LIGNLLSGY  S LF  CTFS NGA MGHKLQ DAAPYKKFAE
DL1 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL2 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL13 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL16 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
ST4 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
R17 replic      PDVLQTSIRS LIGNLLSGYQ SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
J20 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
MS2 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGAPMGHKLQ DAAPYKKFAE
fr replica      PDVLQTSIRS LIGNLLSGYS SQLFRHCTFS NGASMGHKLQ DAAPYKKFAE
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
Consensus       QATVTPRAL  AA LV DQC  PWIRH      ESY FRLV G NGVFTVPKNN
DL1 replic      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL2 replic      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL13 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL16 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
ST4 replic      QATVTPRALR AALLVRDQCA PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
R17 replic      QATVTPRALR AALLVRDQCA PWIRHAVHYN ESYEFRLVVG NGVFTVPKNN
J20 replic      QATVTPRALR AALLVRDQCA PWIRHAVRYN ESYKFRLVVG NGVFTVPKNN
MS2 replic      QATVTPRALR AALLVRDQCA PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
fr replica      QATVTPRALK AAVLVKDQCS PWIRHSHVFP ESYTFRLVGG NGVFTVPKNN

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
Consensus       KIDRAACKEP DMNMYLQKGV G FIRRRL   VGIDLNDQ I NQ LAQQGS
DL1 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
DL2 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
DL13 repli      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
DL16 repli      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
ST4 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQSI NQLLAQQGSV
R17 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQSI NQRLAQQGSA
J20 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLRS VGIDLNDQTI NQRLAQQGSV
MS2 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQSI NQRLAQQGSV
fr replica      KIDRAACKEP DMNMYLQKGV GGFIRRRLKT VGIDLNDQTI NQRLAQQGSR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus       DGSLATIDLS SASDSISDRL VW FLPPELY SYLD IRSHY G   G  IRW
DL1 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
DL2 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
DL13 repli      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
DL16 repli      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
ST4 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
R17 replic      DGSLATIDLS SASDSISDRL VWNFLPPELY SYLDRIRSHY GIVDGETIRW
J20 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIIDGETIRW
MS2 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
fr replica      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDMIRSHY GYVNGKMIRW

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       ELFSTMGNGF TFELESMIFW AIV ATQIHF  N GTIGIYG DDIICP EIA
DL1 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
DL2 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
DL13 repli      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
DL16 repli      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
ST4 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
R17 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
J20 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
MS2 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
fr replica      ELFSTMGNGF TFELESMIFW AIVRATQIHF RNTGTIGIYG DDIICPTEIA
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus       PRVLEAL  Y GFKPNLRKTF  SG FRESC  AH  RGVDVK PFYI KP
DL1 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
DL2 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
DL13 repli      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
DL16 repli      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
ST4 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
R17 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCS AHFYRGVDVK PFYIKKPVDN
J20 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIRKPVDN
MS2 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
fr replica      PRVLEALSFY GFKPNLRKTF TSGSFRESCG AHYFRGVDVK PFYIKKPITD


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus       LF LMLI NR  RGWGVV G   DPRLY VW   LS  VP    FGGTDL ADY
DL1 replic      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
DL2 replic      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
DL13 repli      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
DL16 repli      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
ST4 replic      LFALMLILNR LRGWGVVGGM SDPRLYKVWV RLSSQVPSMF FGGTDLAADY
R17 replic      LFALMLILNR LRGWGVVGGM SDPRLYKVWV RLSSQVPSMF FGGTDLAADY
J20 replic      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
MS2 replic      LFALMLILNR LRGWGVVGGM SDPRLYKVWV RLSSQVPSMF FGGTDLAADY
fr replica      LFSLMLILNR IRGWGVVNGI ADPRLYEVWE KLSRLVPRYL FGGTDLQADY


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        460        470        480        490        500
Consensus       YVVSPP      Y K    R   A  RT GF  LAR A  RK  FS KH SGRY
DL1 replic      YVVSPPTAVS VYTKTAYGRL LADTRTSGFR LARIAKERKR FSEKHDSGRY
DL2 replic      YVVSPPNAVS VYTKTAYGRL LADARTSGFR LARIAKERKH FSEKHDSGRY
DL13 repli      YVVSPPNAVS VYTKTAYGRL LADARTSGFR LARIAKERKH FSEKHGSGRY
DL16 repli      YVVSPPNAVS VYTKTAYGRL LADARTSGFR LARTAKERKH FSEKHDSGRY
ST4 replic      YVVSPPTAVS VYTKTPYGRL LADTRTSGFR LARIARERKF FSEKHDSGRY
R17 replic      YVVSPPTAVS VYTKTPYGRL LADTRTSGFR LARIARERKF FSEKHDSGRY
J20 replic      YVVSPPTAVS VYTKTAYGRL LADARTSGFR LARIAKERKH FSEKHDSGRY
MS2 replic      YVVSPPTAVS VYTKTPYGRL LADTRTSGFR LARIARERKF FSEKHDSGRY
fr replica      YVVSPPILKG IYSKMNGRRE YAEARTTGFK LARIARWRKH FSDKHDSGRY


                ....|....| ....|....| ....|....| ....|....| ....|
                        510        520        530        540
Consensus       IAWFHTGGE  TDSMKSAGVR   RTSEWL P VP FPQECGP ASSP
DL1 replic      IAWFHTGGEI TDSMKSAGVR VMRTSEWLTP VPTFPQECGP ASSPR
DL2 replic      IAWFHTGGEI TDSMKSAGVR VIRTSEWLTP VPIFPQECGP ASSPR
DL13 repli      IAWFHTGGEI TDSMKSAGVR VIRTSEWLTP VPIFPQECGP ASSPR
DL16 repli      IAWFHTGGEI TDSMKSAGVR VIRTSEWLTP VPIFPQECGP ASSPR
ST4 replic      IAWFHTGGEV TDSMKSAGVR IMRTSEWLTP VPTFPQECGP ASSPR
R17 replic      IAWFHTGGEI TDSMKSAGVR IMRTSEWLTP VPTFPQECGP ASSPR
J20 replic      IAWFHTGGEI TDSMKSAGVR VMRTSEWLTP VPTFPQECGP ASSPR
MS2 replic      IAWFHTGGEI TDSMKSAGVR VIRTSEWLTP VPTFPQECGP ASSPR
fr replica      IAWFHTGGEI TDSMKSAGVR VMRTSEWLQP VPVFPQECGP ASSPQ
```

Appendix B2
Group II

Amino acid sequences of *Leviviridae* Group II. Note the YGDD motif in all *Leviviridae* replicase proteins.

Group II isolates DL10, DL20, T72, GA, KUI and TL2. Partial genome sequences of strain TL2 were available in GenBank. The most conserved stretch of amino acids are highlighted in red. Gaps indicate degenerative sites.

Group II maturation protein:

Alignment: Align Group II maturation.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MFPK NIDR  Y V L SYD  KG   SDDSF E  ENYL QN RS  YKPGY
DL10 matur      MFPKSNIDRN YKVTLISYDE KGKLVSDDSF EQVENYLFQN RSNTYKPGYI
DL20 matur      MFPKSNIDRN YKVKLISYDK KGNLVSDDSF EQVENYLFQN RSNTYKPGYI
GA maturat      MFPKSNIDRN YKVKLISYDK KGKLVSDDSF EQVENYLFQN RSTTYKPGYI
T72 matura      MFPKQNIDRT YKVTLISYDK KGNVTSDDSF ESTENYLLQN RSNSYKPGYV
KU1 matura      MFPKQNIDRI YHVKLVSYDN KGKVTSDDSF ESVENYLLQN RSTTYKPGYI

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       R  FR PTNF WNG R F QP VG FTRKL   GGRQ ADYGI VNPNKFT NS
DL10 matur      RRDFRRPTNF WNGFRYFNQP VGTFTRKLSD GGRQVADYGI VNPNKFTGNS
DL20 matur      RRDFRRPTNF WNGYRCFNQP VGTFTRKLSD GGRQVADYGI VNPNKFTANS
GA maturat      RRDFRRPTNF WNGYRCFNQP VGTFTRKLSD GGRQVADYGI VNPNKFTANS
T72 matura      RKGFRKPTNF WNGYRYFNQP VGVFTRKLDN GGRQVADYGI VNPNKFTANS
KU1 matura      RKDFRKPTNF WNGYRYFHQP VGTFTRKLSD GGRQEADYGI VNPNKFTANS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       QHLG NMVIY PGPFSINID  RASVEVLNKL SQSNLNIGVA IAEAKMTASL
DL10 matur      QHLGDNMVIY PGPFSINIDH RASVEVLNKL SQSNLNIGVA IAEAKMTASL
DL20 matur      QHLGDNMVIY PGPFSINIDQ RASVEVLNKL SQSNLNIGVA IAEAKMTASL
GA maturat      QHLGDNMVIY PGPFSINIDQ RASVEVLNKL SQSNLNIGVA IAEAKMTASL
T72 matura      QHLGENMVIY PGPFSINIDN RASVEVLNKL SQSNLNIGVA IAEAKMTASL
KU1 matura      QHLGENMVIY PGPFSINIDN RASVEVLNKL SQSNLNIGVA IAEAKMTASL
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       L  QSI LIR AYTAAKRG W REV SQLLI  EHRF  P  D LGGRWLELQY
DL10 matur      LAKQSIALIR AYTAAKRGKW REVLSQLLIS EHRFRAPVKD LGGRWLELQY
DL20 matur      LAKQSIALIR AYTAAKRGNW REVFSQLLIS EHRFRAPAKD LGGRWLELQY
GA maturat      LAKQSIALIR AYTAAKRGNW REVLSQLLIS EHRFRAPAKD LGGRWLELQY
T72 matura      LSRQSISLIR AYTAAKRGKW REVLSQLLIA EHRFTRPSRD LGGRWLELQY
KU1 matura      LSRQSIALIR AYTAAKRGKW REVLSQLLIA EHRFTRPSKD LGGRWLELQY

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       GWLPLMSD K A YDLLTQT  LPA MPLRV  RTVG THNYK VRNVESAGDT
DL10 matur      GWLPLMSDMK AAYDLLTQTK LPAFMPLRVT RTVGGTHNYK VRNVESAGDT
DL20 matur      GWLPLMSDLK AAYDLLTQTK LPAFMPLRVT RTVGGTHNYK VRNVESAGDT
GA maturat      GWLPLMSDLK AAYDLLTQTK LPAFMPLRVT RTVGGTHNYK VRNVESAGDT
T72 matura      GWLPLMSDIK AGYDLLTQTK LPALMPLRVT RTVGGTHNYK VRNVESAGDT
KU1 matura      GWLPLMSDIK AGYDLLTQTH LPAFMPLRVS RTVGATHNYK VRNVESAGDT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       WSY  RLSVN YRIWYFISDP RLAWASSLGL LNPLEIYWEK TPWSFVVDWF
DL10 matur      WSYSDRLSVN YRIWYFISDP RLAWASSLGL LNPLEIYWEK TPWSFVVDWF
DL20 matur      WSYSDRLSVN YRIWYFISDP RLAWASSLGL LNPLEIYWEK TPWSFVVDWF
GA maturat      WSYRHRLSVN YRIWYFISDP RLAWASSLGL LNPLEIYWEK TPWSFVVDWF
T72 matura      WSYSDRLSVN YRIWYFISDP RLAWASSLGL LNPLEIYWEK TPWSFVVDWF
KU1 matura      WSYSDRLSVN YRIWYFISDP RLAWASSLGL LNPLEIYWEK TPWSFVVDWF

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       LPVGNLIEAM SNPLGLDIIS GTKTWQLESK  NA   A GW  GTAKL AYA
DL10 matur      LPVGNLIEAM SNPLGLDIIS GTKTWQLESK LNATLTASGW SGTAKLSAYA
DL20 matur      LPVGNLIEAM SNPLGLDIIS GTKTWQLESK LNATLPAPGW SGTAKLTAYA
GA maturat      LPVGNLIEAM SNPLGLDIIS GTKTWQLESK LNATLPASGW SGTAKLTAYA
T72 matura      LPVGNLIEAM SNPLGLDIIS GTKTWQLESK LNATINASGW SGTAKLSAYA
KU1 matura      LPVGNLIEAM SNPLGLDIIS GTKTWQLESK MNASLKADGW VGTAKLSAYA

                ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390
Consensus          DRSTFYS FPTP PYVKS PLSGLH ANA LALINQRLKR
DL10 matur      KAYDRSTFYS FPTPMPYVKS PLSGLHLANA LALINQRLKR
DL20 matur      KAYDRSTFYS FPTPLPYVKS PLSGLHLANA LALINQRLKR
GA maturat      KAYDRSTFYS FPTPLPYVKS PLSGLHLANA LALINQRLKR
T72 matura      KAYDRSTFYS FPTPMPYVKS PLSGLHMANA LALINQRLKR
KU1 matura      NNGDRSTFYS FPTPLPYVKS PLSGLHMANA LALINQRLKR
```

Group II capsid protein:

Alignment: Align Group II capsid.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA
DL10 coat       MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA
DL20 coat       MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA
GA coat         MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA
T72 coat        MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA
KU1 coat        MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA
TL2  coat       MATLRSFVLV DNGGTGNVTV VPVSNANGVA EWLSNNSRSQ AYRVTASYRA

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       SGADKRKY I KLEVPKIVTQ  VNGVELP S AWKA ASIDL TIPIFAATDD
DL10 coat       SGADKRKYTI KLEVPKIVTQ VVNGVELPVS AWKAYASIDL TIPIFAATDD
DL20 coat       SGADKRKYTI KLEVPKIVTQ VVNGVELPVS AWKAYASIDL TIPIFAATDD
GA coat         SGADKRKYAI KLEVPKIVTQ VVNGVELPGS AWKAYASIDL TIPIFAATDD
T72 coat        SGADKRKYTI KLEVPKIVTQ TVNGVELPVS AWKAYASIDL TIPIFAATDD
KU1 coat        SGADKRKYTI KLEVPKIVTQ SVNGVELPVS AWKAFASIDL TIPIFAATDD
TL2  coat       SGADKRKYTI KLEVPKIVTQ VVNGVELPVS AWKAYASIDL TIPIFAATDD

                ....|....| ....|....| ....|....|
                        110        120        130
Consensus       VT ISKSLAG LFK G P A  AISSQSGFYA
DL10 coat       VTVISKSLAG LFKVGNPIAE AISSQSGFYA
DL20 coat       VTVISKSLAG LFKVGDPIAD AISSQSGFYA
GA coat         VTVISKSLAG LFKVGNPIAE AISSQSGFYA
T72 coat        VTLISKSLAG LFKIGNPVAD AISSQSGFYA
KU1 coat        VTLISKSLAG LFKIGNPVAD AISSQSGFYA
TL2  coat       VTVISKSLAG LFKVGNPIAD AISSQSGFYA
```

Group II lysis protein:

Alignment: Align Group II lysis.

```
             ....|....| ....|....| ....|....| ....|....| ....|....|
                      10         20         30         40         50
Consensus           G    K      S  S   K                   D    F
T72 lysis    MPSLHRVGST PKAFFSIGSE SNTGKPMFRF TEIKKTLCMD RTRDCAVRFH
KU1 lysis    MPSLHRVGST PKAFFSIGSE SITGKPMFRF TEIEKTLCMD RTRDCAVRFH
DL10 lysis   ------MGLK AKHKENLCSD SQRSKRLYVW IALA--IVLS ---DFTSIFS
DL20 lysis   ------MGLK AKHKENLCSD SQRSKRLYVW IALA--IVLS ---DFTSIFS
GA lysis     ------MGLK AKHKENLCSD SERSKRLYVW IALA--IVLS ---DFTSIFS
TL2 lysis    ------MGLK AKHKENLCSD SQRSKRLYVW IALA--IVLS ---DFTSIFS


             ....|....| ....|....| ....|....
                      60         70
Consensus         L           D
T72 lysis    VYLQSLDLGS SDPHSPDFDG LAYLR----
KU1 lysis    VYLQSLDLGS SDPHSPDFDG LAYLRDECL
DL10 lysis   HWIWGLLILI LQ-TLMDLPT FVMSV----
DL20 lysis   HWIWGLLILI LR-TLMDLPT FVMNV----
GA lysis     HWIWGLLILY LQ-TLMDLPT FVMNV----
TL2 lysis    HWIWGLLILI LR-TLMDLPT FVMNV----
```

Group II replicase protein:

Alignment: Align Group II replicase.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MFRF EI KT LCMDRTRDCA VRFHVYLQSL D GSSDP SP DFDGLAYLRD
T72             MFRFTEIKKT LCMDRTRDCA VRFHVYLQSL DLGSSDPHSP DFDGLAYLRD
DL10            MFRFTEIEKT LCMDRTRDCA VRFHVYLQSL DMGSSDPHSP DFDGLAYLRD
DL20            MFRFTEIEKT LCMDRTRDCA VRFHVYLQSL DLGSSDPHSP DFDGLAYLRD
GA              MFRFREIEKT LCMDRTRDCA VRFHVYLQSL DLGSSDPLSP DFDGLAYLRD
KU1             MFRFTEIEKT LCMDRTRDCA VRFHVYLQSL DLGSSDPHSP DFDGLAYLRD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       ECLTKHPSLG  SNSDA RKE LAYAKLMDSD QRCKIQNSNG YD SHI   V
T72             ECLTKHPSLG DSNSDALRKE LAYAKLMDSD QRCKIQNSNG YDLSHIDSGV
DL10            ECLTKHPSLG NSNSDARRKE LAYAKLMDSD QRCKIQNSNG YDYSHIESSV
DL20            ECLTKHPSLG DSNSDARRKE LAYAKLMDSD QRCKIQNSNG YDYSHIESGV
GA              ECLTKHPSLG DSNSDARRKE LAYAKLMDSD QRCKIQNSNG YDYSHIESGV
KU1             ECLTKHPSLG DSNSDALRKE LAYAKLMDSD QRCKIQNSNG YDLSHIDAGV

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       L GIL TA A   A LL GFE SHFLNDCSFS NGASQGFKL  DAAPFKKIAG
T72             LNGILLTAKA SIAKLLMGFE SHFLNDCSFS NGASQGFKLQ DAAPFKKIAG
DL10            LSGILKTAQA LVANLLTGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG
DL20            LSGILKTAQA LVANLLTGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG
GA              LSGILKTAQA LVANLLTGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG
KU1             LNGILLTAKA LIAKLLIGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       QATVTAPAY   AV AVKTC  PW  YMQETY GDET WFRRV YGNGLFSVPK
T72             QATVTAPAYD LAVHAVKTCG PWLRYMQETY GDETRWFRRV YGNGLFSVPK
DL10            QATVTAPAYN IAVAAVKTCA PWYAYMQETY GDETRWFRRV YGNGLFSVPK
DL20            QATVTAPAYD IAVAAVKTCA PWYAYMQETY GDETKWFRRV YGNGLFSVPK
GA              QATVTAPAYD IAVAAVKTCA PWYAYMQETY GDETKWFRRV YGNGLFSVPK
KU1             QATVTAPAYD LAVLAVKTCG PWLRYMQETY GDETRWFRRV YGNGLFSVPK

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       NNKIDRAACK EPDMNMYLQK GAGSFIR RL RSV IDLNDQ T NQELARLG
T72             NNKIDRAACK EPDMNMYLQK GAGSFIRRRL RSVNIDLNDQ TRNQELARLG
DL10            NNKIDRAACK EPDMNMYLQK GAGSFIRKRL RSVGIDLNDQ TCNQELARLG
DL20            NNKIDRAACK EPDMNMYLQK GAGSFIRKRL RSVGIDLNDQ TRNQELARLG
GA              NNKIDRAACK EPDMNMYLQK GAGSFIRKRL RSVGIDLNDQ TRNQELARLG
KU1             NNKIDRAACK EPDMNMYLQK GAGSFIRRRL RSVNIDLNDQ TRNQELARLG
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus       SIDGSLATID LSSASDS SD RLVWDLLPPH VYSYL RIR  SFTMIDG LH
T72             SIDGSLATID LSSASDSVSD RLVWDLLPPH VYSYLHRIRS SFTMIDGQLH
DL10            SIDGSLATID LSSASDSVSD RLVWDLLPPH VYSYLARIRS SFTMIDGRLH
DL20            SIDGSLATID LSSASDSISD RLVWDLLPPH VYSYLARIRS SFTMIDGRLH
GA              SIDGSLATID LSSASDSISD RLVWDLLPPH VYSYLARIRT SFTMIDGRLH
KU1             SIDGSLATID LSSASDSVSD RLVWDLLPPH VYSYLHRIRS SFTMIDGRLH

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       KW LFSTMGN GFTFELESMI FWALS   M    GVTG LG  YGDDIIVP
T72             KWNLFSTMGN GFTFELESMI FWALSKSVMS YLGVTGLLGI YGDDIIVPTK
DL10            KWGLFSTMGN GFTFELESMI FWALSKSVML SMGVTGSLGI YGDDIIVPVE
DL20            KWGLFSTMGN GFTFELESMI FWALSKSVML SMGVTGSLGV YGDDIIVPVE
GA              KWGLFSTMGN GFTFELESMI FWALSKSIML SMGVTGSLGI YGDDIIVPVE
KU1             KWNLFSTMGN GFTFELESMI FWALSNTVMS YLGVTGLLGI YGDDIIVPTK

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       C P LL VLS AVNFLPN  K TFTTGYFRES CGAHFFK A   KPFYCKRPM
T72             CAPLLLQVLS AVNFLPNQKK TFTTGYFRES CGAHFFKGAS VKPFYCKRPM
DL10            CAPTLLKVLS AVNFLPNKKK TFTTGYFRES CGAHFFKGAD MKPFYCKRPM
DL20            CAPTLLKVLS AVNFLPNQKK TFTTGYFRES CGAHFFKGAD MKPFYCKRPM
GA              CRPTLLKVLS AVNFLPNEEK TFTTGYFRES CGAHFFKDAD MKPFYCKRPM
KU1             CAPLLLQVLS AVNFLPNQKK TFTTGYFRES CGAHFFKGAS VKPFYCKRPM

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       ETLPD  LLC NRIRGW T G G SDPRLFPI WKEFADMIPP KFKGGCNLDR
T72             ETLPDIMLLC NRIRGWGTIG GISDPRLFPI WKEFADMIPP KFKGGCNLDR
DL10            ETLPDVMLLC NRIRGWQTVG GMSDPRLFPI WKEFADMIPP KFKGGCNLDR
DL20            ETLPDVMLLC NRIRGWQTVG GMSDPRLFPI WKEFADMIPP KFKGGCNLDR
GA              ETLPDVMLLC NRIRGWQTVG GMSDPRLFPI WKEFADMIPP KFKGGCNLDR
KU1             ETLPDVLLLC NRIRGWGTIG GISDPRLFPI WKEFADMIPP KFKGGCNLDR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       DTYLVSPDKP G  LVR A   RSGFN  F     ENGRY HW LHMGSGEV E
T72             DTYLVSPDKP GKTLVRVAKK RSGFNHKFRS DYENGRYIHW LHMGSGEVLE
DL10            DTYLVSPDKP GVTLVRVAKV RSGFNHSFPY GHENGRYVHW LHMGSGEVLE
DL20            DTYLVSPDKP GVTLVRIAKV RSGFNHAFPY GYENGRYVHW LHMGSGEVLE
GA              DTYLVSPDKP GVSLVRIAKV RSGFNHAFPY GHENGRYVHW LHMGSGEVLE
KU1             DTYLVSPDKP GVTLVRDATV RSGFNYKFRR RQENGRYIHW LHMGSGEVSE

                ....|....| ....|....| ....|....| ..
                       510        520        530
Consensus       TISSAR RCK PNSEWR QIP LFPQE EACV LS
T72             TISSARFRCK PNSEWRTQIP LFPQEIEACV LS
DL10            TISSARFRCK PNSEWRTQIP LFPQELEACV LS
DL20            TISSARFRCK PNSEWRTQIP LFPQELEACV LS
```

```
GA          TISSARYRCK PNSEWRTQIP LFPQELEACV LS
KU1         TISSARFRCK PNSEWRIQIP LFPQEVEACV LS
```

Appendix B3
Group III

Amino acid sequences of *Leviviridae* Group III. Note the YGDD motif in all *Leviviridae* replicase proteins.

Group III isolates BR12, BZ1, HL4-9, TW18, VK, QB, MX1 and M11. The most conserved stretch of amino acids are highlighted in red. Gaps indicate degenerative sites.

Group III maturation protein:

Alignment: Align Group III maturation.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
Consensus       MP LPR LRF G   E L DF QELW P      S    P YT   G
BR12 matur      MPKLPRGLRF GADNEILNDF QELWFPDLFI ESSDTHPWYT LKGRVLNAHM
BZ1 matura      MPKLPRGLRF GADNEILNDF QELWFPDLFI ESSDTHPWYT LKGRVLNAHL
HL49 assem      MPKLPRGLRF GADNEILNDF QELWFPDLFI ESSDTHPWYT LKGRVLNAHL
TW18 matur      MPKLPRGLRF GADNEILNDF QELWFPDLFI ESSDTHPWYT LKGRVLNAHL
VK assembl      MPKLPRGLRF GADNEILNDF QELWFPDLFI ESSDTHPWYT LKGRVLNAHI
QB assembl      MPKLPRGLRF GADNEILNDF QELWFPDLFI ESSDTHPWYT LKGRVLNAHL
MX1 assemb      MPRLPRALRF GPNMEVLSDF QELWYPESII DSDVKYPLYT FRGSIGGSFF
M11 assemb      MPRLPRGLRF GANMEVLNDF QELWYPESRV DSDTIFPLYT FKGNMGGSFF

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
Consensus       D    N   R   RRTPH  T VPIAS GLRP  T V YDP   L F     R
BR12 matur      DDRLPNVSGR QVRRTPHRAT VPIASTGLRP VTTVQYDPAA LSF-LLNARV
BZ1 matura      DDRLPNVGGR QIRRTPHRAT VPIASSGLRP VTTVQYDPTA LSF-LLNARV
HL49 assem      DDRLPNVGGR QIRRTPHRAT VPIASSGLRP VTTVQYDPTA LSF-LLNARV
TW18 matur      DDRLPNVGGR QVRRTPHRVT VPIASSGLRP VTTVQYDPAA LSF-LLNARV
VK assembl      DDRLPNVSGR QVRRTPHRVT VPIASSGLRP VTTVQYDPAA LSF-LLNARV
QB assembl      DDRLPNVGGR QVRRTPHRVT VPIASSGLRP VTTVQYDPAA LSF-LLNARV
MX1 assemb      DSYGTNNIVR EIRRTPHCAT VPIASSGLRP CTSVWYDPTS LLFRIPEMRA
M11 assemb      DTYGTNNIVR QVRRTPHRAT VPIASSGLRP CTSVWYDPSS LLFRIPEMRA
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
Consensus        WD G GD     V  DF F  T APK FDFS NSL PRY  A FSAFNAKYG
BR12 matur      DWDFGNGDSA DLVIKDFVFR TFAPKDFDFS NSLAPRYTQA FSAFNAKYGV
BZ1 matura      DWDFGDGDSA DLVINDFLFR TFAPKEFDFS NSLAPRYTQA FSAFNAKYGT
HL49 assem      DWDFGNGDSA NLVINDFLFR TFAPKEFDFS NSLVPRYTQA FSAFNAKYGT
TW18 matur      DWDFGNGDSA NLVINDFLFR TSAPKEFDFS NSLVPRYTQA FSAFNAKYGT
VK assembl      DWDFGNGDSA DLVIKDFVFR TFAPKDFDFS NSLVPRYTQA FSAFNAKYGV
QB assembl      DWDFGNGDSA NLVINDFLFR TFAPKEFDFS NSLVPRYTQA FSAFNAKYGT
MX1 assemb      EWDNGMGDAG DIVYKDFLFS TPAPKEFDFS NSLAPRYSNA FSAFNAKYGV
M11 assemb      VWDNGMGDTG DIVYNDFLFN TPAPKEFDFS NSLAPRYSNA FSAFNAKYGV

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
Consensus        IGEG ET K Y  LLLRRL     RAV  GD LR LR    S Y  G WKP T
BR12 matur      MIGEGLETIK YLGLLLRRLR EGYRAVKRGD LRALRRVIQS YHKGKWKPTT
BZ1 matura      MIGEGLETIK YLGLLLRRLR EGFRAVKHGD LRALRRVIQS YHKGKWKPTT
HL49 assem      MIGEGLETIK YLGLLLRRLR EGYRAVKRGD LRALRRVIQS YHKGKWKPTT
TW18 matur      MIGEGLETIK YLGLLLRRLR EGYRAVKRGD LRALRRVIQS YHNGKWKPTT
VK assembl      MIGEGLETIK YLGLLLRRLR EGYRAVKRGD LRALRRVIQS YHKGKWKPTT
QB assembl      MIGEGLETIK YLGLLLRRLR EGYRAVKRGD LRALRRVIQS YHNGKWKPAT
MX1 assemb      IIGEGHETLK YFALLLRRLH KAVRAVRHGD LRGLRKILDS YNKGRWKPAT
M11 assemb      IIGEGRETLK YFALLLRRLH KAVRAVRHGD LRGLRRILDS YHKGHWKPAT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
Consensus       AGNLWLEFRY GL PLF DI   VM DW     D IQ   RFS VGHGED
BR12 matur      AGNLWLEFRY GLMPLFYDIR DVMLDWQSRH DKIQRLLRFS VGHGEDYVVK
BZ1 matura      AGNLWLEFRY GLMPLFYDIR DVMLDWQNRH DKIQRLLRFS VGHGEDYVVE
HL49 assem      AGNLWLEFRY GLMPLFYDIK DVMLDWQKRH DRIQRLLRFS VGHGEDYVVK
TW18 matur      AGNLWLEFRY GLMPLFYDIK DVMLDWQNRH DKIQRLLRFS VGHGEDCVVE
VK assembl      AGNLWLEFRY GLMPLFYDIR DVMLDWQNRH DKIQRLLRFS VGHGEDYVVK
QB assembl      AGNLWLEFRY GLMPLFYDIR DVMLDWQNRH DKIQRLLRFS VGHGEDYVVE
MX1 assemb      AGNLWLEFRY GLTPLFHDIK SVMDDWNRIN DKIQKLRRFS VGHGEDFKLS
M11 assemb      AGNLWLEFRY GLVPLFHDIK DVMNDWTRIN DKIQKYRRFS VGHGEDFKLS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus        D LYP     F L GEIT   RRHR GI YA NR GYA FDN GS RPVSDWK
BR12 matur      FDNLYPALAY FKLKGEITLE RRHRHGISYA NREGYAVFDN GSLRPVSDWK
BZ1 matura      FDNLYPAVAY FKLKGEITLE RRHRHGISYA NREGYAVFDN GSLRPVSDWK
HL49 assem      FDSLYPAVSY FKLRGEITLE RRHRHGISYA NREGYAVFDN GSLRPVSDWK
TW18 matur      FDNLYPAVAY FKLKGEITLE RRHRHGISYA NREGYAVFDN GSLRPVSDWK
VK assembl      FDNLYPALAY FKLKGEITLE RRHRHGISYA NRGGYAVFDN GSLRPVSDWK
QB assembl      FDNLYPAVAY FKLKGEITLE RRHRHGISYA NREGYAVFDN GSLRPVSDWK
MX1 assemb      IDGLYPGLTH FRLSGEITVQ RRHRWGITYA NREGYATFDN GSIRPVSDWK
M11 assemb      IDGLYPGLTH FRLSGEITVQ RRHRWGIVYA NREGYATFDN GSIRPVSDWK
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       ELA AFINP  EVAWELTPYS F  DWF NVG DI  QQ Q Y  NI IVDG
BR12 matur      ELAVAFINPH EVAWELTPYS FVADWFLNVG DILAQQGQLY HNIEIVDGFD
BZ1 matura      ELATAFINPQ EVAWELTPYS FVVDWFLNVG DILAQQGQLY HNIDIVDGFD
HL49 assem      ELATAFINPH EVAWELTPYS FVVDWFLNVG DILAQQGQLY HNIDIVDGFD
TW18 matur      ELATAFINPH EVAWELTPYS FVVDWFLNVG DILAQQGQLY HNIDIVDGFD
VK assembl      ELAVAFINPH EVAWELTPYS FVVDWFLNVG DILAQQGQLY HNIDIVDGFD
QB assembl      ELATAFINPH EVAWELTPYS FVVDWFLNVG DILAQQGQLY HNIDIVDGFD
MX1 assemb      ELANAFINPG EVAWELTPYS FIVDWFINVG DIIEQQKQWY QNIDIVDGYQ
M11 assemb      ELANAFINPG EVAWELTPYS FVVDWFINVG DIIEQQKQLY QNIDIVDGYQ

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       RRDIR  S    KG RNG PV  V       D  FY R H     P ATL  DT
BR12 matur      RRDIRFKSFT IKGERNGQPV NVSADLSTID TFYSRLHASN IPFATLDLDT
BZ1 matura      RRDIRFKSFT IKGERNGQPV NVSADLSAVD TFYSRLHTSS IPFATLDLDT
HL49 assem      RRDIRLKSFT IKGERNGRPV NVSADLSAVD SFYSRLHTTK LPFATLDLDT
TW18 matur      RRDIRLKSFT IKGERNGRPV NVSADLSAVD LFYSRLHTSS LPFATLDLDT
VK assembl      RRDIRFKSFT IKGERNGQPV NVSADLSAID TFYSRLHASN IPFATLDLDT
QB assembl      RRDIRLKSFT IKGERNGRPV NVSASLSAVD LFYSRLHTSN LPFATLDLDT
MX1 assemb      RRDIRMRSVS LKGVRNGIPV RVTGSVELVD SFYNRSHTTR IPQATLAIDT
M11 assemb      RRDIRMRSVT LKGVRNGIPV RVTGSVELVD SFYNRSHTTR IPQATLELDT

                ....|....| ....|....| .
                       410        420
Consensus        FSS KHV D S  L TQ  K R
BR12 matur      TFSSFKHVLD SIFLLTQRIK R
BZ1 matura      TFSSFKHVLD SIFLLTQRIK R
HL49 assem      TFSSYKHVLD SIFLLTQRIK R
TW18 matur      TFSSYKHVLD SIFLLTQRIK R
VK assembl      TFSSFKHVLD SVALLTQRIK R
QB assembl      TFSSFKHVLD SIFLLTQRVK R
MX1 assemb      SFSSIKHVMD SISLITQRIK R
M11 assemb      SFSSIKHVLD SISLITQXIK R
```

Group III capsid protein:

Alignment: Align Group III capsid.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MAKL   TL   IGK G  TL  LNPRGVNPT NGVA LS AG AVPALEKRVT
BR12 coat       MAKLETVTLS NIGKDGKQTL VLNPRGVNPT NGVASLSEAG AVPALEKRVT
BZ1 capsid      MAKLETVTLS NIGKDGQKTL NLNPRGVNPT NGVASLSEAG AVPALEKRVT
HL49 capsi      MAKLETVTLS NIGKDGKQTL VLNPRGVNPT NGVAALSQAG AVPALEKRVT
TW18 capsi      MAKLETVTLS NIGKDGQQTL VLNPRGVNPT NGVAALSQAG AVPALEKRVT
VK capsid       MAKLETVTLS NIGKDGKQTL VLNPRGVNPT NGVASLSEAG AVPALEKRVT
QB coat         MAKLETVTLG NIGKDGKQTL VLNPRGVNPT NGVASLSQAG AVPALEKRVT
MX1 capsid      MAKLQAITLS GIGKNGDVTL NLNPRGVNPT NGVAALSEAG AVPALEKRVT
M11 capsid      MAKLQAITLS GIGKKGDVTL DLNPRGVNPT NGVAALSEAG AVPALEKRVT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus        SVSQPSRNR KNYKVQVKIQ NPT C A G  CDPSVTR AY  DVTFSFTQY
BR12 coat       VSVSQPSRNR KNYKVQVKIQ NPTACPANGS CDPSVTRQAY ADVTFSFTQY
BZ1 capsid      ISVSQPSRNR KNYKVQVKIQ NPTACSANGS CDPSVTRQAY ADVTFSFTQY
HL49 capsi      VSVSQPSRNR KNYKVQVKIQ NPTACTANGS CDPSVTRQAY ADVTFSFTQY
TW18 capsi      VSVSQPSRNR KNYKVQVKIQ NPTACTANGS CDPSVTRQAY ADVTFSFTQY
VK capsid       VSVSQPSRNR KNYKVQVKIQ NPTACSANGS CDPSVTRQAY ADVTFSFTQY
QB coat         VSVSQPSRNR KNYKVQVKIQ NPTACTANGS CDPSVTRQAY ADVTFSFTQY
MX1 capsid      ISVSQPSRNR KNYKVQVKIQ NPTSCTASGT CDPSVTRSAY ADVTFSFTQY
M11 capsid      ISVSQPSRNR KNYKVQVKIQ NPTSCTASGT CDPSVTRSAY SDVTFSFTQY

                ....|....| ....|....| ....|....| ...
                        110        120        130
Consensus       ST EERA VR TEL ALLA P  L  AID LN PAY
BR12 coat       STDEERAFVR TELAALLAGP LLIDAIDRLN PAY
BZ1 capsid      STDEERAFVR TELAALLADP LLIDAIDQLN PAY
HL49 capsi      STDEERAFVR TELVALLASP LLIDAIDQLN PAY
TW18 capsi      STDEERAFVR TELIALLASP LLIDAIDQLN PAY
VK capsid       STDEERAFVR TELAALLAGP LLIDAIDRLN PAY
QB coat         STDEERAFVR TELAALLASP LLIDAIDQLN PAY
MX1 capsid      STDEERALVR TELKALLADP MLIDAIDNLN PAY
M11 capsid      STVEERALVR TELQALLADP MLVNAIDNLN PAY
```

Group III read-through protein:

Alignment: Align Group III read-through protein.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MAKL   TL   IGK G  TL  LNPRGVNPT NGVA LS AG AVPALEKRVT
TW18 readt      MAKLETVTLS NIGKDGQQTL VLNPRGVNPT NGVAALSQAG AVPALEKRVT
HL49 readt      MAKLETVTLS NIGKDGKQTL VLNPRGVNPT NGVAALSQAG AVPALEKRVT
BR12 readt      MAKLETVTLS NIGKDGKQTL VLNPRGVNPT NGVASLSEAG AVPALEKRVT
BZ1 readth      MAKLETVTLS NIGKDGQKTL NLNPRGVNPT NGVASLSEAG AVPALEKRVT
VK readthr      MAKLETVTLS NIGKDGKQTL VLNPRGVNPT NGVASLSEAG AVPALEKRVT
QB readthr      MAKLETVTLG NIGKDGKQTL VLNPRGVNPT NGVASLSQAG AVPALEKRVT
M11 readth      MAKLQAITLS GIGKKGDVTL DLNPRGVNPT NGVAALSEAG AVPALEKRVT
MX1 readth      MAKLQAITLS GIGKNGDVTL NLNPRGVNPT NGVAALSEAG AVPALEKRVT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus        SVSQPSRNR KNYKVQVKIQ NPT C A G  CDPSVTR AY  DVTFSFTQY
TW18 readt      VSVSQPSRNR KNYKVQVKIQ NPTACTANGS CDPSVTRQAY ADVTFSFTQY
HL49 readt      VSVSQPSRNR KNYKVQVKIQ NPTACTANGS CDPSVTRQAY ADVTFSFTQY
BR12 readt      VSVSQPSRNR KNYKVQVKIQ NPTACPANGS CDPSVTRQAY ADVTFSFTQY
BZ1 readth      ISVSQPSRNR KNYKVQVKIQ NPTACSANGS CDPSVTRQAY ADVTFSFTQY
VK readthr      VSVSQPSRNR KNYKVQVKIQ NPTACSANGS CDPSVTRQAY ADVTFSFTQY
QB readthr      VSVSQPSRNR KNYKVQVKIQ NPTACTANGS CDPSVTRQAY ADVTFSFTQY
M11 readth      ISVSQPSRNR KNYKVQVKIQ NPTSCTASGT CDPSVTRSAY SDVTFSFTQY
MX1 readth      ISVSQPSRNR KNYKVQVKIQ NPTSCTASGT CDPSVTRSAY ADVTFSFTQY

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       ST EERA VR TEL ALLA P  L  AID LN PAYT L   G  SG  P
TW18 readt      STDEERAFVR TELIALLASP LLIDAIDQLN PAYTLLIAGG GSGSNPDS-V
HL49 readt      STDEERAFVR TELVALLASP LLIDAIDQLN PAYTLLIAGG GSGSKPDPVI
BR12 readt      STDEERAFVR TELAALLAGP LLIDAIDRLN PAYTLLIAGG GSGENPDP-V
BZ1 readth      STDEERAFVR TELAALLADP LLIDAIDQLN PAYTLLIAGG GSGENPDP-V
VK readthr      STDEERAFVR TELAALLAGP LLIDAIDRLN PAYTLLIAGG GSGENPDP-V
QB readthr      STDEERAFVR TELAALLASP LLIDAIDQLN PAYTLLIAGG GSGSKPDP-V
M11 readth      STVEERALVR TELQALLADP MLVNAIDNLN PAYTALLGVG -SGPSPGPGP
MX1 readth      STDEERALVR TELKALLADP MLIDAIDNLN PAYTALLGDG -SGPSPVPGP

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus        PDPP  PPP G G YTCPF  IW L   YE        W IY  A EL  R F
TW18 readt      IPDPPIDPPP GTGKYTCPFA IWSLEEVYEP PTRNRPWPIY NAVELRPRKF
HL49 readt      IPDPPIDPPP GTGKYTCPFA IWSLEEVYEP PTKNRPWPIY NAIELQPRKF
BR12 readt      DPDPPIDPPP GTGSYTCPFA IWSLEEVYEP PTIDRPWPIY RAVELSSRKF
BZ1 readth      DPDPPIDPPP GSGSYTCPFA IWSLEEVYEP PTSDRPWPIY YAVELQPRDF
VK readthr      NPDPPIDPPP GTGTYTCPFA IWSLEEVYEP PTSDRPWPIY HAVELPPRDF
QB readthr      IPDPPIDPPP GTGKYTCPFA IWSLEEVYEP PTKNRPWPIY NAVELQPREF
M11 readth      DPDPPPEPPP GTGSYTCPFR IWDLSSVYEA ANSSHSWDIY NAVELSPRNF
```

MX1 readth NPDPPLEPPP GTGSYTCPFR IWDLSSIYEA ANSSHSWDIY NAVELSPRKF

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       DV L DLLGN T WRDWD RL  YTTFRG R  NGYIDLDA   L  D
TW18 readt      DVALEDLLGN TNWRDWDSRL SYTTFRGCRG NGYIDLDATY LATDQAMRDQ
HL49 readt      DVALKDLLGN TKWRDWDSRL SYTTFRGCRS NGYIDLDASY LATDQAMRDQ
BR12 readt      DVALDDLLGN TKWRDWDSRL RYTTFRGCRG NGYIDLDATY LATDQAMLDQ
BZ1 readth      DVALDDLLGN TKWRDWDSRL SYTTFRGCRG NGYIDLDATY LATDQAMLDQ
VK readthr      DVALDDLLGN TEWRDWDSRL RYTTFRGCRG NGYIDLDATY LATDQAMLDQ
QB readthr      DVALKDLLGN TKWRDWDSRL SYTTFRGCRG NGYIDLDATY LATDQAMRDQ
M11 readth      DVALDDLLGN TNWRDWDGRL RYTTFRGCRG NGYIDLDATS LMKDEYLTSS
MX1 readth      DVTLDDLLGN TDWRDWDGRL RYTTFRGSRG NGYIDLDATS LMQDEYLTSS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       KY  R GK P G FG  E F  YLKSINAYCS LSDI AY  D GV VGFWRDP
TW18 readt      KYDIREGKKP GAFGNIERFI YLKSINAYCS LSDIAAYHAD GVIVGFWRDP
HL49 readt      KYDIREGKKP GAFGNVERFI YLKSINAYCS LSDIAAYHAD GVIVGFWRDP
BR12 readt      KYDIRTGKRP GAFGKIEQFI YLKSINAYCS LSDIAAYHAD GVIVGFWRDP
BZ1 readth      KYDIRTGKKP GAFGNIEQFI YLKSINAYCS LSDIAAYHAD GVIVGFWRDP
VK readthr      KYDIRTGKRP GAFGNIEQFI YLKSINAYCS LSDIAAYHAD GVIVGFWRDP
QB readthr      KYDIREGKKP GAFGNIERFI YLKSINAYCS LSDIAAYHAD GVIVGFWRDP
M11 readth      KYLVREGKRP GVFGNIERFV YLKSINAYCS LSDITAYRTD GVIVGFWRDP
MX1 readth      KYLVREGKRP GAFGSIERFV YLKSINAYCS LSDITAYHSD GVVVGFWRDP

                ....|....| ....|....| ....|....
                        310        320
Consensus       SSGGAIPFDF   FD  KCPI QAVIVVPR
TW18 readt      SSGGAIPFDF TKFDKTKCPI QAVIVVPRA
HL49 readt      SSGGAIPFDF TKFDKTKCPI QAVIVVPRA
BR12 readt      SSGGAIPFDF TQFDKTKCPI QAVIVVPRA
BZ1 readth      SSGGAIPFDF TEFDKTKCPI QAVIVVPRA
VK readthr      SSGGAIPFDF TEFDKTKCPI QAVIVVPRA
QB readthr      SSGGAIPFDF TKFDKTKCPI QAVIVVPRA
M11 readth      SSGGAIPFDF NEFDSNKCPI QAVIVVPRL
MX1 readth      SSGGAIPFDF SEFDSNKCPI QAVIVVPRL
```

Group III replicase:

Alignment: Align Group III replicase.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus          MSKT  S   SLS  LRR AANTRI VE  NLALSIANDL   A     F
BR12 repli      MKCMSKTASS HNSLSAQLRR AANTRIEVEG NLALSIANDL MLAYGQSPFS
BZ1 replic      MKCMSKTASS RNSLSAQLRR AANTRIEVEG NLALSIANDL MLAYGQSPFN
HL49 repli      MKCMSKTASS HNSLSAQLRR AANTRIEVEG NLALSIANDL LLAYGQSPFN
TW18 repli      MKCMSKTASS HNSLSAQLRR AANTRIEVEG NLALSIANDL LLAYGQSPFN
VK replica      MKCMSKTASS HNSLSAQLRR AANTRIEVEG NLALSIANDL MLAYGQSPFS
QB replica      MKCMSKTASS RNSLSAQLRR AANTRIEVEG NLALSIANDL LLAYGQSPFN
MX1 replic      ---MSKTLQS RKSLSGKLRR AANTRIVVEG NLALSIANDL LSALDVEPFN
M11 replic      ---MSKTSQS RKSLSGKLRR AANTRIVVED NLALSIANDL LSALDVESFS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       SE  CIS  P  F    D FR   YL AE MS KYD FSLGI  TEA AW KFL
BR12 repli      SESECISLSP KFDGTPDNFR INYLKAEIMS KYDDFSLGID TEAAAWKKFL
BZ1 replic      SESECISLGP KFDETPDNFR INYLKAEIMS KYDDFSLGID TEAAAWEKFL
HL49 repli      SEAECISLSP RFDGTPDHFR INYLKAEVMS KYDDFSLGID TEAAAWEKFL
TW18 repli      SEAECISLSP RFDGTPDDFR INYLKAEVMS KYDDFSLGID TEAAAWEKFL
VK replica      SESECISLSP KFDGTPDNFR INYLKAEIMS KYDDFSLGID TEAAAWEKFL
QB replica      SEAECISFSP RFDGTPDDFR INYLKAEIMS KYDDFSLGID TEAVAWEKFL
MX1 replic      SEEDCISRSP KFGISPDQFR NSYLRAEIMS KYDSFSLGIN TEAVAWEKFL
M11 replic      SEEDCISRSP KFDLSADQFR NSYLAAEIMS KYDSFSLGIN TEAVAWEKFL

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       AAEAECA TN  RLYRP Y E DFNFSLGE C  HMARRKI K L GD    E
BR12 repli      AAEAECALTN ARLYRPNYSE DFNFSLGESC LHMARRKIVK LIGDAPSVEG
BZ1 replic      AAEAECALTN ARLYRPNYSE DFNFSLGESC IHMARRKIVK LIGDAPSVEG
HL49 repli      AAEAECALTN ARLYRPDYSE DFNFSLGESC IHMARRKIVK LIGDAPSVEG
TW18 repli      AAEAECALTN ARLYRPDYSE DFNFSLGESC IHMARRKIAK LIGDAPSVEG
VK replica      AAEAECALTN ARLYRPNYSE DFNFSLGESC LHMARRKIVK LIGDAPSVEG
QB replica      AAEAECALTN ARLYRPDYSE DFNFSLGESC IHMARRKIAK LIGDVPSVEG
MX1 replic      AAEAECAKTN LRLYRPNYNE DFNFSLGETC IHMARRKIVK LLGDSVPFEA
M11 replic      AAEAECAITN QRLYRPNYNE DFNFSLGEAC IHMARRKIVK LLGDSVPFEA

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus        LRHCRFSGG ATTTN R  G HPSFKFAL Q  CTPRA  YV  AL A T  D
BR12 repli      MLRHCRFSGG ATTTNNRSYG HPSFKFALPQ ACTPRALKYV LALRASTHFD
BZ1 replic      MLRHCRFSGG ATTTNNRSYG HPSFKFALPQ ACTPRALKYV LALRASTHFD
HL49 repli      MLRHCRFSGG ATTTNNRSHG HPSFKFALPQ ACTPRALKYV LALRASTHFD
TW18 repli      MLRHCRFSGG ATTTNNRSYG HPSFKFALPQ ACTPRALKYV LALRASTHFD
VK replica      MLRHCRFSGG ATTTNNRSYG HPSFKFALPQ ACTPRALKYV LALRASTHFD
QB replica      MLRHCRFSGG ATTTNNRSYG HPSFKFALPQ ACTPRALKYV LALRASTHFD
MX1 replic      VLRHCRFSGG ATTTNSRLYG HPSFKFALAQ ECTPRAVPYV QALKACTNMD
```

M11 replic MLRHCRFSGG ATTTNNRSYG HPSFKFALTQ ECTPRAVPYV QALKACTGMD

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus             SPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG G  R  L  W
BR12 repli      IRVSDISPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GILRDRLRCW
BZ1 replic      IRVSDISPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GILRDRLRCW
HL49 repli      IRVSDISPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GILRDRLRCW
TW18 repli      IRISDISPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GILRDRLRCW
VK replica      IRVSDISPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GILRDRLRCW
QB replica      IRISDISPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GILRDRLRCW
MX1 replic      LGITKVSPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GVIREKLHLW
M11 replic      LGITKVSPFN KAVTVPKNSK TDRCIAIEPG WNMFFQLGIG GVIREKLRLW

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus        IDLNDQTIN Q RA  GS     LATVDLS  ASD ISLAL  ELL PP WF
BR12 repli      GIDLNDQTIN QHRAHEGSVT NDLATVDLSA ASDSISLALC ELLLPPGWFE
BZ1 replic      GIDLNDQTIN QRRAHEGSVT NNLATVDLSA ASDSISLALC ELLLPPGWFE
HL49 repli      GIDLNDQTIN QHRAHEGSVT NNLATVDLSA ASDSISLALC ELLLPPGWFE
TW18 repli      GIDLNDQTIN QRRAHEGSVT NNLATVDLSA ASDSISLALC ELLLPPGWFE
VK replica      GIDLNDQTIN QRRAHEGSVT NDLATVDLSA ASDSISLALC ELLLPPGWFE
QB replica      GIDLNDQTIN QRRAHEGSVT NNLATVDLSA ASDSISLALC ELLLPPGWFE
MX1 replic      NIDLNDQTIN QVRAYSGSCS NELATVDLSS ASDTISLALV ELLLPPAWFK
M11 replic      GIDLNDQTIN QTRAYLGSRD DNLATVDLSR ASDTISLALV ELLMPPEWFK

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       VL  LRS  G  LP G    Y EKISSMGNG  TFELESLIFA  LARS CE L
BR12 repli      VLTDLRSPKG QLPDGSVITY EKISSMGNGY TFELESLIFA SLARSVCEIL
BZ1 replic      VLMDLRSPKG RLPNGSVVTY EKISSMGNGY TFELESLIFA SLARSVCEIL
HL49 repli      VLMDLRSPKG RLPDGSVVTY EKISSMGNGY TFELESLIFA SLARSVCEIL
TW18 repli      VLMDLRSPKG RLPDGSVVTY EKISSMGNGY TFELESLIFA SLARSVCEIL
VK replica      VLMDLRSPKG RLPDGSVIIY EKISSMGNGY TFELESLIFA SLARSVCEIL
QB replica      VLMDLRSPKG RLPDGSVVTY EKISSMGNGY TFELESLIFA SLARSVCEIL
MX1 replic      VLTDLRSRRG MLPDGRIITY EKISSMGNGF TFELESLIFA ALARSLCELL
M11 replic      VLLALRSPKG ILPDGTVITY EKISSMGNGY TFELESLIFA ALARSLCELL

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus        L  S VTVY GDDIILPS A    L EVF Y VGF TN KKT F  GPFRESC
BR12 repli      DLDSSEVTVY GDDIILPSRA VPALQEVFKY VGFTTNTKKT FSEGPFRESC
BZ1 replic      DLDSSEVTVY GDDIILPSCA VPALQEVFKY VGFTTNTKKT FSEGPFRESC
HL49 repli      DLDSSEVTVY GDDIILPSCA VPALREVFKY VGFTTNTKKT FSEGPFRESC
TW18 repli      DLDSSEVTVY GDDIILPSCA VPALREVFKY VGFTTNTKKT FSEGPFRESC
VK replica      NLDSSEVTVY GDDIILPSRA VPALQEVFKY VGFTTNTKKT FSEGPFRESC
QB replica      DLDSSEVTVY GDDIILPSCA VPALREVFKY VGFTTNTKKT FSEGPFRESC
MX1 replic      NLQPSSVTVY GDDIILPSDA CSSLIEVFSY VGFRTNEKKT FFDGPFRESC
M11 replic      GLRPSDVTVY GDDIILPSDA CSPLVEVFSY VGFRTNKKKT FSSGPFRESC
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       GKHY  GVDV TPFYIR RIV  P DLILVLN   YRWATIDG VWDPR   VY
BR12 repli      GKHYYSGVDV TPFYIRRRIV SPADLILVLN NLYRWATIDG VWDPRAHSVY
BZ1 replic      GKHYYSGVDV TPFYIRRRIV SPADLILVLN NLYRWATIDG VWDPRAHSVY
HL49 repli      GKHYYSGVDV TPFYIRHRIV NPTDLILVLN NLYRWATIDG VWDPRAHSVY
TW18 repli      GKHYYSGVDV TPFYIRHRIV TPADLILVLN NLYRWATIDG VWDPRAHPVY
VK replica      GKHYYSGVDV TPFYIRRRIV SPADLILVLN NLYRWATIDG VWDPRAHSVY
QB replica      GKHYYSGVDV TPFYIRHRIV SPADLILVLN NLYRWATIDG VWDPRAHSVY
MX1 replic      GKHYFMGVDV TPFYIRHRIV SPSDLILVLN QMYRWATIDG VWDPRVYPVY
M11 replic      GKHYFLGVDV TPFYIRRRIV SPSDLILVLN QMYRWATIDG VWDPRVYPVY


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus        KYR  LP   L RN  PDGY GDGALVGSVL   PFA NRGW  R VP I D
BR12 repli      LKYRKLLPKQ LQRNTIPDGY GDGALVGSVL INPFAKNRGW IRYVPVIIDH
BZ1 replic      LKYRKLLPKQ LQRNTIPDGY GDGALVGSVL INPFAKNRGW IRYVPVIIDH
HL49 repli      LKYRKLLPKQ LQRNTIPDGY GDGALVGSVL INPFAKNRGW IRYVPVIIDH
TW18 repli      LKYRKLLPKQ LQRNTIPDGY GDGALVGSVL INPFAKNRGW IRYVPVITDH
VK replica      LKYRKLLPKQ LQRNTIPDGY GDGALVGSVL INPFAKNRGW IRYVPVIIDH
QB replica      LKYRKLLPKQ LQRNTIPDGY GDGALVGSVL INPFAKNRGW IRYVPVITDH
MX1 replic      TKYRRLLPDI LRRNVVPDGY GDGALVGSVL TSPFAENRGW VRRVPMIIDK
M11 replic      TKYRRYLPEI LRRNVVPDGY GDGALVGSVL ISPFAENRGW VRRVPMIIDK


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus         DR R E G SYLY L S     E     P   G
BR12 repli      TRDRERTESG SYLYDLFSRC FSESNDGLPL RGPSSCDPVY PLAIDQLICK
BZ1 replic      TRDRERTEPG SYLYDLFSRC FSESNDGLPL RGPSSCDSIY RFAVDQLICK
HL49 repli      TRDRERVESG SYLYDLFSRC FSEGNDGLPL RGPSSCDSVD LSAIDQLICR
TW18 repli      TRDRERAELG SYLYDLFSRC LPESNDGLPL RGPSSCDSVN LSAVDQLICR
VK replica      TRDRERTESG SYLYDLFSRY FSESNDGLPL RGPSSCDSIY PLAIDQLICK
QB replica      TRDRERAELG SYLYDLFSRC LSESNDGLPL RGPSGCDSAD LFAIDQLICR
MX1 replic      KKDRVRDERG SYLYELWSLQ QLECDSEFPF NGSLVVGTND GVCTYRHRER
M11 replic      RKDRVRDEYG SYLYELWSLQ QLECDSEFPF NGSLVVGSTD GTLAYAHRER


                ....|....| ....|....| ....|....| ....|....| ..
                       560        570        580        590
Consensus          T  S       D   I C  SRVLAPYG  F       L
BR12 repli      SNPTKISRST GKFDVQYIAC SSRVLAPYGV FQGTKVASLH EA
BZ1 replic      SNPTKVSRST GKFDIQYIAC SSRVLAPYGV FQGTKVMPLH EA
HL49 repli      SNPTKISRST GKFDIQYIAC GSRVLAPYGV FQGTKVTSLH EV
TW18 repli      SNPTKISRST GKFDIQYIAC SSRVLAPYGV FQGTKVTSLH EV
VK replica      SNPTKISRST GKFDIQYIAC SSRVLAPYGV FQGTKVASLH EA
QB replica      SNPTKISRST GKFDIQYIAC SSRVLAPYGV FQGTKVASLH EA
MX1 replic      -VSTAISDSV GAYDIVWIPC SSRVLAPYGD FRRHEGSILK --
M11 replic      -LPTVISDAV SAFDIMWIPC SSRVLAPYGD FRRHEGSILK MG
```

Appendix B4
Group IV

Amino acid sequences of *Leviviridae* Group IV. Note the YGDD motif in all *Leviviridae* replicase proteins.

Group IV isolates BR1, BR8, HB-P22, HB-P24, SP, NL95 and FI. The most conserved stretch of amino acids are highlighted in red. Gaps indicate degenerative sites.

Group IV maturation protein:

Alignment: Align Group IV maturation.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MP LP GLRF GS GE  NDF   LWFP          G   L  GY
BR1 assemb      MPTLPRGLRF GSNGEVLNDF EALWFPERHT VDLSNGTCKL TGYITNLPGY
BR8 assemb      MPTLPRGLRF GSNGEVLNDF EALWFPERHT VDLSNGTCKL TGYITNLPDY
SP assembl      MPTLPRGLRF GSNGEVLNDF EALWFPERHT VDLSNGTCKL TGYITNLPGY
HB-P22 asm      MPALPRGLRF GSNGEILNDF NELWFPELVS SELNLGTYNL TGYISNLPGY
HB-P24 asm      MPTLPRGLRF GSNGEVVNDF NALWFPERET FDSGLGSYEL TGYVSNQPGY
NL95 assem      MPTLPRGLRF GSNGEIVNDF NALWFPEREA FDLELGSYTL TGYVSNQPGY
FI assembl      MPTLPIGLRF GSKGEILNDF SALWFPKRVS FDSQLGRYEL SGYLNGQ--F

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus             N        TP R TV PVNH GYRPV TTVEY P GT   RLDG V
BR1 assemb      SNIFPNKGVT VARTPYRSTV PVNHLGYRPV TTVEYIPDGT YVRLDGHVKF
BR8 assemb      SNTFPNKGVT VARTPYRSTV PVNHLGYRPV TTVEYIPDGT YVRLDGHVKF
SP assembl      SDIFPNKGVT AARTPYRSTV PVNHLGYRPV TTVEYIPDGT YVRLDGHVKF
HB-P22 asm      EVKERNKGSH VLRTPYRSTV PVNHLGYRPV TTVEYNPLGT FIRLDGDVKF
HB-P24 asm      ETRMRNPSMH CIRTPHRSTV PVNHLGYRPV TTVEYTPNGT FIRLDGDVKF
NL95 assem      TTRMRNPRMH CVRTPHRSTV PVNHFGYRPV TTVEYVPNGT FIRLDGDVKF
FI assembl      SDYGRNPALQ VSKTPHRATV PVNHLGYRPV TTVEYVPNGT FVRLDGTVRI

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus        G  V      L N    LA  Q GFDYQSVI GPRFS  F A FSTKYG LLG
BR1 assemb      EGGLVNGSVD LTNYVISLAA QGGFDYQSVI GPRFSAHFSA FSTKYGVLLG
BR8 assemb      EGGLVNGSVD LTNFVISLAA QGGFDYQSVI GPRFSAHFSA FSTKYGVLLG
SP assembl      EGDLVNGSVD LTNFVISLAA QGGFDYQSVI GPRFSARFSA FSTKYGVLLG
HB-P22 asm      SGGLVSASLR LDNYVVGLAS QAGFDYQSVI GPRFSSQFSA FSTKYGTLLG
HB-P24 asm      SGGAVSGSLK LSNYVVNLAA QGGFDYQSVI GPRFSAQFAA FSTKYGTLLG
NL95 assem      SGGSVSGSLK LNNFVVNLAS QGGFDYQSVI GPRFSSQFSA FSTKYGALLG
FI assembl      SGELVNGTVR LDNYIVNLAA QGGFDYQSVI GPRFSSQFSA FSTKYGTLLG
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
Consensus       EGRETL YLL L  RR REG   AV  GD KR LRN     EP     G
BR1 assemb      EGRETLKYLL LLLRRMREGY RAVRRGDLKR LRNVISTFEP TSSKGRKARS
BR8 assemb      EGRETLKYLL LLLRRMREGY RAVRRGDFKR LRNVISTFEP ISPRGKMARS
SP assembl      EGRETLKYLL LVVRRMREGY RAVRRGDLKR LRNVISTFEP STIKGKRARA
HB-P22 asm      EGRETLSYLL LLFRRVREGI RAVRRGDLKR LRNLVRSYEP TSVNGRRQAA
HB-P24 asm      EGRETLSYLL LLFRRVREGI RAVKHGDLKR LRNLIRTFEP RSVPGKRQRA
NL95 assem      EGRETLSYLL LLFRRMREGI RAVRRGDLKR LRNLIRTFEP RSISGRRQRT
FI assembl      EGRETLSYLL LLFRRMREGF LAVKRGDLKR LRNVIRTFEP RSQAGKRQRS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
Consensus        F   Y D                 W  SSAS  LWLEFRYGL MPLFYDI S
BR1 assemb      EFSQTYRDKL TGNKVEVKPS EGKWNSSSAS DLWLEFRYGL MPLFYDIQSV
BR8 assemb      EFSQTYRDKL TGNKVKVKPS EGKWNSSSAS DLWLEFRYGL MPLFYDIQSV
SP assembl      EFSQTYRDKL TGNKVEVRPS EGKWNSSSAS DLWLEFRYGL MPLFYDIQSV
HB-P22 asm      SFSGTYRDEL AN-------- -GKWKDSSAS DLWLEFRYGL MPLFYDIQSL
HB-P24 asm      SFSEAYRDRT AN-------- -GSWKDSSAS DLWLEFRYGL MPLFYDIQSV
NL95 assem      TFSETYRDRL VN-------- -DGWKDSSAS DLWLEFRYGL MPLFYDIKSV
FI assembl      SFNKSYCDKL AS-------- -GEWKNSSAS NLWLEFRYGL MPLFYDIQSV

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus       MEDFMR HK  IAK QRFSAG HGKL  V    F P   F  E VTAVLQRRHR
BR1 assemb      MEDFMRVHKK IAKIQRFSAG HGKLETVSSR FYPDVHFALE VTAVLQRRHR
BR8 assemb      MEDFMRVHKK IAKIQRFSAG HGKLETVSSR FYPDVHFALE VTAVLQRRHR
SP assembl      MEDFMRVHKK IAKIQRFSAG HGKLETVSSR FYPDVHFSLE VTAVLQRRHR
HB-P22 asm      MEDFMRVHKK IAKIQRFSAG HGKLVEVSGT FYPDVHFGLE VTAVLQRRHR
HB-P24 asm      MEDFMRVHKK IAKLQRFSAG HGKLVEVKGK FFPDPHFALE VTAVLQRRHR
NL95 assem      MEDFMRIHKK IAKLQRFSAG HGKLVTVKGR FFPDPHFAIE VTAVLQRRHR
FI assembl      MEDFMRVHKR IAKIQRFSAG HGKLEKVSDI FYPSTYFQLE VTAVLQRRHR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       WGV YQDT      F NG L  PV DW TAA  A LNPAE AW E TP SFV D
BR1 assemb      WGVIYQDTGS YATFNNGRLV PVKDWKTAAF ALLNPAEVAW EVTPYSFVVD
BR8 assemb      WGVVYQDTGS YATFNNGRLV PVKDWKTAAF ALLNPAEVAW EITPYSFVVD
SP assembl      WGVIYQDTGS FATFNNGRLV PVKDWKTAAF ALLNPAEVAW EVTPYSFVVD
HB-P22 asm      WGVIYQDTGS YATFDNGRLV PVKDWQTAAF ALLNPAETAW ELTPYSFVAD
HB-P24 asm      WGVIYQDTGS FATFNNGRLI PVRDWQTAAL ALLNPAETAW ELTPYSFVAD
NL95 assem      WGVIYQDTGS LPPFNNGQLI PVRDWQTAAL ALLNPAETAW ELTPYSFVAD
FI assembl      WGVIYQDTDT FATFDNGRLI PVRDWQTAAL AFLNPAETAW ELTPLSFVAD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       WFVNVGDMLE Q    L   V DVVDGFDRKD   L S SVRV      A
BR1 assemb      WFVNVGDMLE QMG-QLYRHV DVVDGFDRKD IKLKSVSVRV LTGDSAHVAK
BR8 assemb      WFVNVGDMLE QMG-QLYRHV DVVDGFDRKD IRLKSISVRV LTSDSAHVAS
SP assembl      WFVNVGDMLE QMG-QLYRHV DVVDGFDRKD IKLKSVSVRV LTNDVAHVAS
HB-P22 asm      WFVNVGDMLE QIG-QLYRHV DVVDGFDRKD VKLRSVSVRV IA-NGATNTQ
HB-P24 asm      WFVNVGDMLE QIG-QLYRHV DVVDGFDRKD VKLKSVSVRV IR-DDATHTS
```

```
NL95 assem   WFVNVGDMLE QTRPALSVNV DVVDGFDRKD VKLRSVSVRV IR-DDATTTS
FI assembl   WFVNVGDMLE QIG-QLYRYV DVVDGFDRKD VKLKSVSVRV IA-PSADSAT
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       F L  A LLH   YSRVHTVA FPQISPQ D  E RS KHVID S ALLTQR K
BR1 assemb      FTLRQARLLH SYYSRVHTVA FPQISPQLDT EIRSVKHVID SIALLTQRVK
BR8 assemb      FTLRQARLLH SYYSRVHTVA FPQISPQLDT EIRSVKHVID SIALLTQRVK
SP assembl      FQLRQAKLLH SYYSRVHTVA FPQISPQLDT EIRSVKHVID SIALLTQRVK
HB-P22 ass      FVLRRASLLH SFYSRVHTVA FPQISPQLDT EVRSVKHVID SIALLTQRFK
HB-P24 ass      FSLAGAVLLH SFYSRVHTVA FPQISPQLDT EIRSVKHVID SIALLTQRFK
NL95 assem      FNLSGAKLLH SFYSRVHTVA FPQISPQLDT EVRSFKHVID SVALLTQRFK
FI assembl      FELRRAELLH GFYSRVHTVA FPQISPQIDA EVRSVKHVID SLALLTQRFK


                ..

Consensus
BR1 assemb      R-
BR8 assemb      R-
SP assembl      R-
HB-P22 asm      R-
HB-P24 asm      RR
NL95 assem      KR
FI assembl      R-
```

Group IV capsid protein:

Alignment: Align Group IV capsid.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MAKLN VTL    GK    T  TLTPRGVNPT NGVA LSEAG AVPALEKRVT
BR1 coat        MAKLNQVTLS KIGKNGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
BR8 coat        MAKLNQVTLS KLGKNGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
SP coat         MAKLNQVTLS KIGKNGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
HB-P22 coa      MAKLNQVTLT KLGKAGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
HB-P24 coa      MAKLNKVTLT GIGKAGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
NL95 coat       MAKLNKVTLT GIGKAGNQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
FI coat         MAKLNTVTLT KLGKEANKTM TLTPRGVNPT NGVATLSEAG AVPALEKRVT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       VSVAQPSRNR KN K QIKLQ NPTACT DA  DPSVTRS    D TLSFTSYS
BR1 coat        VSVAQPSRNR KNFKVQIKLQ NPTACTKDA- DPSVTRSAFA DVTLSFTSYS
BR8 coat        VSVAQPSRNR KNFKVQIKLQ NPTACTKDAC DPSVTRSAFA DVTLSFTSYS
SP coat         VSVAQPSRNR KNFKVQIKLQ NPTACTRDAC DPSVTRSAFA DVTLSFTSYS
HB-P22 coa      VSVAQPSRNR KNYKVQIKLQ NPTACTKDAC DPSVTRSAFA DLTLSFTSYS
HB-P24 coa      VSVAQPSRNR KNYKIQIKLQ NPTACTKDAC DPSVTRSAFA DLTLSFTSYS
NL95 coat       VSVAQPSRNR KNYKVQIKLQ NPTACTKDAC DPSVTRSGSR DVTLSFTSYS
FI coat         VSVAQPSRNR KNFKVQIKLQ NPTACTKDAC DPSVTRSAFA DLTLSFTSYS

                ....|....| ....|....| ....|....| ..
                        110        120        130
Consensus       T  ERAL RT ELAALL D L I DAIDNLNP AY
BR1 coat        TDEERALIRT ELAALLQDNL IIDAIDNLNP AY
BR8 coat        TDAERALIRT ELAALLQDPL IVDAIDNLNP AY
SP coat         TDEERALIRT ELAALLADPL IVDAIDNLNP AY
HB-P22 coa      TDAERALIRT ELAALLQDPL IVDAIDNLNP AY
HB-P24 coa      TDVERALVRT ELAALLKDDL IVDAIDNLNP AY
NL95 coat       TERERALIRT ELAALLKDDL IVDAIDNLNP AY
FI coat         TDEERALIRT ELAALLADPL ITDAIDNLNP AY
```

Group IV read-through protein:

```
Alignment: Align Group IV read-through.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MAKLN VTL    GK    T  TLTPRGVNPT NGVA LSEAG AVPALEKRVT
BR1 readth      MAKLNQVTLS KIGKNGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
BR8 readth      MAKLNQVTLS KLGKNGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
SP readthr      MAKLNQVTLS KIGKNGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
HB-P22 rea      MAKLNQVTLT KLGKAGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
HB-P24 rea      MAKLNKVTLT GIGKAGDQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
NL95 readt      MAKLNKVTLT GIGKAGNQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
FI readthr      MAKLNTVTLT KLGKEANKTM TLTPRGVNPT NGVATLSEAG AVPALEKRVT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       VSVAQPSRNR KN K QIKLQ NPTACT DAC DPSVTRS    D TLSFTSYS
BR1 readth      VSVAQPSRNR KNFKVQIKLQ NPTACTKDAC DPSVTRSAFA DVTLSFTSYS
BR8 readth      VSVAQPSRNR KNFKVQIKLQ NPTACTKDAC DPSVTRSAFA DVTLSFTSYS
SP readthr      VSVAQPSRNR KNFKVQIKLQ NPTACTRDAC DPSVTRSAFA DVTLSFTSYS
HB-P22 rea      VSVAQPSRNR KNYKVQIKLQ NPTACTKDAC DPSVTRSAFA DLTLSFTSYS
HB-P24 rea      VSVAQPSRNR KNYKIQIKLQ NPTACTKDAC DPSVTRSAFA DLTLSFTSYS
NL95 readt      VSVAQPSRNR KNYKVQIKLQ NPTACTKDAC DPSVTRSGSR DVTLSFTSYS
FI readthr      VSVAQPSRNR KNFKVQIKLQ NPTACTKDAC DPSVTRSAFA DLTLSFTSYS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       T  ERAL RT ELAALL D L I DAIDNLNP AY AALL  S   G   P  P
BR1 readth      TDEERALIRT ELAALLQDNL IIDAIDNLNP AY-AALLVAS SGGGDNPSMP
BR8 readth      TDAERALIRT ELAALLQDPL IVDAIDNLNP AY-AALLVAS SGGGDNPSKP
SP readthr      TDEERALIRT ELAALLADPL IVDAIDNLNP AY-AALLVAS SGGGDNPSDP
HB-P22 rea      TDAERALIRT ELAALLQDPL IVDAIDNLNP AY-AALLAAS SSGGDNP-KP
HB-P24 rea      TDVERALVRT ELAALLKDDL IVDAIDNLNP AY-AALLAAS PGGGNNP-NP
NL95 readt      TERERALIRT ELAALLKDDL IVDAIDNLNP AY-AALLAAS PGGGNNP-YP
FI readthr      TDEERALIRT ELAALLADPL ITDAIDNLNP AYWAALLVSS SGGEVKPPIP

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus        VP  P VKP P GTG   CP F CYR           G     Y  G E  V
BR1 readth      DVPDLPDVKP PDGTGRYTCP FSCYRLGSIY EEGKDGSPDI YERGDEVSVT
BR8 readth      DVPDLPDVKP PDGTGRYTCP FACYRLGSIY EVGKDGSPDI YERGDEVSVT
SP readthr      DVPVVPDVKP PDGTGRYKCP FACYRLGSIY EVGKEGSPDI YERGDEVSVT
HB-P22 rea      DVPDRPDVKP PGGTGSYRCP FTCYRLGNII EVGQNGSPDI YARGDEVQVM
HB-P24 rea      GVPDSPNVKP PGGTGTYRCP FACYRRGELI TEAKDGACAL YALGSEAIVE
NL95 readt      GVPDSPNVKP PGGTGTYRCP FACYRRGELI TEAKDGACAL YACGSEALVE
FI readthr      DVPDVPSVKP PGGTGSFTCP FSCYRLDTII EAGKDGVPDL YEQGPEVTVT
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       F YA EDFLG N  WRNWD R LS YD    R RCRGNGY DL  A  MQ D
BR1 readth      FDYALEDFLG NTNWRNWDQR LSSYDLANRR RCRGNGYIDL DATVMQSDEF
BR8 readth      FDYALEDFLG NTNWRNWDQR LSSYDLANRR RCRGNGYIDL NATAMQSDEC
SP readthr      FDYALEDFLG NTNWRNWDQR LSDYDIANRR RCRGNGYIDL DATAMQSDDF
HB-P22 rea      FDYALEDFLG NTNWRNWDQR LSNYDIANRR RCRGNGYVDL DATAMQTDSF
HB-P24 rea      FDYALEDFLG NVFWRNWDGR LSTYDIDTHR RCRGNGYVDL DATMMQSDAY
NL95 readt      FEYALEDFLG NEFWRNWDGR LSKYDIETHR RCRGNGYVDL DASVMQSDEY
FI readthr      FDYAVEDFLG NTNWRNWDSR LSNYDIGNLR RCRGNGYVDL DATAMQSDSY

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       VLSG Y V K    P  F     Y L          DL  VTAY  SYGMVIGFW
BR1 readth      VLSGRYPVRK VKFPGAFGSI KYLLNIQGDA WLDLSEVTAY RSYGMVIGFW
BR8 readth      VLSGRYPVRK VKFPGAFGSI KYLLNIQGDA WLDLSEVTAY RSYGMVIGFW
SP readthr      VLSGRYGVRK VKFPGAFGSI KYLLNIQGDA WLDLSEVTAY RSYGMVIGFW
HB-P22 rea      VLSGKYPVRK VKFPGAFGAL KYLLNIKDDA WVDLSEVTAY RSYGMVIGFW
HB-P24 rea      VLSGAYDVVK MQPPSIFDSP RYYLHLMDGI YVDLAEVTAY HSYGMVIGFW
NL95 readt      VLSGAYDVVK MQPPGTFDSP RYYLHLMDGI YVDLAEVTAY RSYGMVIGFW
FI readthr      VLSGKYRVRK GLPPGIFASP RYYLELQDGA WVDLAAVTAY RSYGMVIGFW

                ....|....| ....|....| ....|....| ...
                        310        320        330
Consensus       TDSKSPQ P  DFT F    C PVQTVI IPS L
BR1 readth      TDSKSPQLPT DFTQFNSANC PVQTVIIIPS L--
BR8 readth      TDSKSPQLPT DFTQFNSANC PVQTVIIIPS L--
SP readthr      TDSKSPQLPT DFTQFNSANC PVQTVIIIPS L--
HB-P22 rea      TDSKSPQLPS DFTQFDSTNC PVQTVIVIPS L--
HB-P24 rea      TDSKSPQLPT DFTRFNGRNC PVQTVIVIPS L--
NL95 readt      TDSKSPQLPT DFTRFNRHNC PVQTVIVIPS LAT
FI readthr      TDSKSPQIPN DFTRFDSTKC PVQTVIVIPS LD-
```

Group IV replicase:

Alignment: Align Group IV replicase.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       M KTA      IT  L KVDI  FEDDIH SI ANDL  A G    L  AE CI
BR1 replic      MSKTASRRRE ITQLLGKVDI NFEDDIHMSI ANDLFEAYGI PKLQNAEECI
BR8 replic      MSKTASRRRE ITQLLGKVDI NFEDDIHMSI ANDLCEAYGI PMLRDAEECI
SP replica      MPKTASRRRE ITQLLGKVDI NFEDDIHMSI ANDLFEAYGI PKLDSAEECI
HB-P22 rep      MSKTASHRKK ITHLLSKVDI DFEDDIHMSI ANDLFKAFGV APLTSAEQCI
HB-P24 rep      MSKTASHKKK ITQALSKVDI NFEDDIHMSI ANDLFEAYGI APLASAEQCI
NL95 repli      MSKTACHKKK ITQTLSKVDI NFEDDIHMSI ANDLLQACGV APLASAEQCI
FI replica      MSKTASRRRE ITHLLGKVDI SFEDDIHLSI ANDLFRAYGV GELSSAEECI

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus        T FP       D FR  YLR  EILSK   H PLG  TE  A WEKFLAAEEG
BR1 replic      NTAFPSLDQG ADTFRVEYLR AEILSKFDGH PLGTDTEKTA WEKFLAAEEG
BR8 replic      NTAFPSLDQS ADTFRVEYLR AEILSKFDGH PLGINTEEAA WEKFLAAEEG
SP replica      NTAFPSLDQG VDTFRVEYLR AEILSKFDGH PLGIDTEAAA WEKFLAAEEG
HB-P22 rep      NTPFPDTSMT ADAFRIAYLR SEILSKYSAH PLGIDTEAVA WEKFLAAEEG
HB-P24 rep      STPFPDTSMD ADAFRIHYLR SEILSKFSAH PLGIDTEAAA WEKFLAAEEG
NL95 repli      NTPFPDTSMN PDEFRIQYLR SEILSKFSAH PLGIDTEAAA WEKFLAAEEG
FI replica      NTAFPRLDQS PDTFRTEYLR SEILSKFNAH PLGIDTEAVA WEKFLAAEEG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       CR TN RL    KYHDNSILS WGERVIHTAR RKILKLIGE   P GDVALR
BR1 replic      CRLTNERLSQ VKYHDNSILS WGERVIHTAR RKILKLIGES VPFGDVALRC
BR8 replic      CRRTNERLSQ AKYHDNSILS WGERVIHTAR RKILKLIGES VPFGDVALRC
SP replica      CRQTNERLSL VKYHDNSILS WGERVIHTAR RKILKLIGES VPFGDVALRC
HB-P22 rep      CRQTNERLSQ AKYHDNSILS WGERVIHTAR RKILKLIGES VPLGDVALRC
HB-P24 rep      CRQTNERLTK VKYHDNSILS WGERVIHTAR RKILKLIGET VPLGDVALRA
NL95 repli      CRLTNERLTK VKYHDNSILS WGERVIHTAR RKILKLIGEA APLGDVALRA
FI replica      CRLTNARLSS CKYHDNSILS WGERVIHTAR RKILKLIGES VPFGDVALRC

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       RFSGGATTSV NRLHGHPSWK HACPQDVTKR A KYL A K  ACGD   LR
BR1 replic      RFSGGATTSV NRLHGHPSWK HACPQDVTKR AFKYLQAFKR ACGDVVDLRV
BR8 replic      RFSGGATTSV NRLHGHPSWK HACPQDVTKR ALKYLQAFKR ACGDVVDLRV
SP replica      RFSGGATTSV NRLHGHPSWK HACPQDVTKR AFKYLQAFKR ACGDVVDLRV
HB-P22 rep      RFSGGATTSV NRLHGHPSWK HACPQDVTKR AFKYLQAYKM ACGDIVDLRV
HB-P24 rep      RFSGGATTSV NRLHGHPSWK HACPQDVTKR ALKYLMAYKK ACGDVVDLRV
NL95 repli      RFSGGATTSV NRLHGHPSWK HACPQDVTKR ALKYLIAYKK ACGDVVDLRV
FI replica      RFSGGATTSV NRLHGHPSWK HACPQDVTKR ALKYLLAYKK ACGDTDELRI
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus        EVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLW ID
BR1 replic      NEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWKID
BR8 replic      NEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWKID
SP replica      NEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWKID
HB-P22 rep      NEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWGID
HB-P24 rep      NEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWQID
NL95 repli      NEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWHID
FI replica      GEVRTSNKAV TVPKNSKTDR CIAIEPGWNM FFQLGVGAVL RDRLRLWSID

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       LNDQS NQRL ARD S L HL AT DLSAASD SIS  LVELL  PP W   LT
BR1 replic      LNDQSTNQRL ARDGSLLNHL ATIDLSAASD SISLKLVELL LPPEWYDLLT
BR8 replic      LNDQSTNQRL ARDGSLLNHL ATIDLSAASD SISLKLVELL LPPEWYDLLT
SP replica      LNDQSTNQRL ARDGSLLNHL ATIDLSAASD SISLKLVELL MPPEWYDLLT
HB-P22 rep      LNDQSINQRL ARDASQLDHL ATVDLSAASD SISLKLVELL LPPDWFGVLT
HB-P24 rep      LNDQSVNQRL ARDASQLDHL ATVDLSAASD SISLRLVELL MPPAWFDLLT
NL95 repli      LNDQSVNQRL ARDASQLDHL ATVDLSAASD SISLRLVELL MPPAWFDLLT
FI replica      LNDQSLNQRL ARDASQLDHL ATVDLSAASD SISIKLVELL LPPAWFELLT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       DLRSD G LP  G  VTYEKI SSMGNGYTFE LESLIFAA A RSVCELL  D
BR1 replic      DLRSDEGVLP DGRVVTYEKI SSMGNGYTFE LESLIFAALA RSVCELLEID
BR8 replic      DLRSDQGVLP DGRVVTYEKI SSMGNGYTFE LESLIFAALA RSVCELLEID
SP replica      DLRSDEGILP DGRVVTYEKI SSMGNGYTFE LESLIFAAIA RSVCELLEID
HB-P22 rep      DLRSDQGILP DGRAVTYEKI SSMGNGYTFE LESLIFAAIA RSVCELLDLD
HB-P24 rep      DLRSDQGVLP DGRVVTYEKI SSMGNGYTFE LESLIFAALA RSVCELLDLD
NL95 repli      DLRSDQGILP DGRVVTYEKI SSMGNGYTFE LESLIFAALA RSVCELLDLD
FI replica      DLRSDQGVLP NGEVVTYEKI SSMGNGYTFE LESLIFAAIA RSVCELLDLD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus       QS VSVYGDD III   AA   LM VFEYVGF T N KKTF   GPFRESCGKH
BR1 replic      QSTVSVYGDD IIIDTRAAAP LMDVFEYVGF TPNKKKTFCD GPFRESCGKH
BR8 replic      QSTVSVYGDD IIIDTRAAAP LMDVFEYVGF TPNKKKTFCD GPFRESCGKH
SP replica      QSTVSVYGDD IIIDTRAAAP LMDVFEYVGF TPNRKKTFCD GPFRESCGKH
HB-P22 rep      QSTVSVYGDD IIIDSRAAAT LMDVFEYVGF TPNRKKTFVS GPFRESCGKH
HB-P24 rep      QSTVSVYGDD IIIDSRAADV LMAVFEYVGF TPNRKKTFIK GPFRESCGKH
NL95 repli      QSTVSVYGDD IIIDSRAADV LMAVFEYVGF TPNRKKTFIK GPFRESCGKH
FI replica      QSAVSVYGDD IIIPSDAAQT LMDVFEYVGF TANRKKTFIT GPFRESCGKH

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus       W  GVDVTPF YIRRPIR L  DMILVLN  Y RWGT DG WD PR L VY KY
BR1 replic      WFQGVDVTPF YIRRPIRCLA DMILVLNSIY RWGTVDGVWD PRALTVYEKY
BR8 replic      WFQGVDVTPF YIRRPIRCLA DMILVLNSIY RWGTVDGVWD PRALTVYEKY
SP replica      WFQGVDVTPF YIRRPIRCLA DMILVLNSIY RWGTVDGIWD PRALTVYEKY
HB-P22 rep      WHSGVDVTPF YIRRPIRCLV DMILVLNSIY RWGTIDGVWD PRVLPVYQKY
HB-P24 rep      WHSGVDVTPF YIRRPIRCLA DMILVLNSIY RWGTIDGVWD PRVLPVYQKY
```

NL95 repli WHSGVDVTPF YIRRPIRCLA DMILVLNSIY RWGTIDGVWD PRVLPVYQKY
FI replica WFLGVDVTPF YIRRPIRSLA DMILVLNNLY RWGTVDGVWD PRALTVYQKY

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        460        470        480        490        500
Consensus          LPR WRR N IPDGYGDG ALVG A TNP FV V N  R  YPVLVEVQ D
BR1 replic      LKLLPRNWRR NRIPDGYGDG ALVGLATTNP FVIVKNYSRL YPVLVEVQRD
BR8 replic      LKLLPRNWRR NRIPDGYGDG ALVGLATTNP FVIVKNYSRL YPVLVEVQRD
SP replica      LKLLPRNWRR NRIPDGYGDG ALVGLATTNP FVIVKNYSRL YPVLVEVQRD
HB-P22 rep      VNMLPRNWRR NTIPDGYGDG ALVGLATTNP FVIVKNFSRL YPVLVEVQKD
HB-P24 rep      VKLLPRDWRR NTIPDGYGDG ALVGLATTNP FVIVRNYSRW YPVLVEVQRD
NL95 repli      VKLLPRDWRR NTIPDGYGDG ALVGLATTNP FVIVRNYSRW YPVLVEVQRD
FI replica      VKLLPRNWRR NTIPDGYGDG ALVGSALTNP FVLVRNFQRE YPVLVEVQKD


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        510        520        530        540        550
Consensus         R E G YL Y  LR R  R   PFL   D    FDE PLAT  LRRKTGRYK
BR1 replic      VKRSEVGSYL YALLRDRETR YSPFLRDADR TGFDEAPLAT SLRRKTGRYK
BR8 replic      VKRSEVGSYL YSLLRNRETR YSPFLRDADR TGFDEAPLAT SLRRKTGRYK
SP replica      VKRSEEGSYL YALLRDRETR YSPFLRDADR TGFDEAPLAT SLRRKTGRYK
HB-P22 rep      VKRHEYGSYL YAMLRDRETR YNPFLRVADG TGFDEAPLAT SLRRKTGRYK
HB-P24 rep      AKRHEFGSYL YALLRDREAR YNPFLRTADG SGFDETPLAT SLRRKTGRYK
NL95 repli      AKRHEFGSYL YALLRDRDAR YSPFLRVADG TGFDEAPLAT SLRRKTGRYK
FI replica      TPRSEKGAYL YHLLRDREAR HNPFLYDTDW VRFDEAPLAT RLRRKTGRYK


                ....|....| ....|....| ....|....| ....|..
                        560        570        580
Consensus       VAWIQDSAFI RPPY   G P EVK A
BR1 replic      VAWIQDSAFI RPPYFITGIP EVKLAS---- -------
BR8 replic      VAWIQDSAFI RPPYFITGIP EVKLAS---- -------
SP replica      VAWIQDSAFI RPPYLITGIP EVKLAS---- -------
HB-P22 rep      VAWIQDSAFI RPPYFLTGLP EVKLAS---- -------
HB-P24 rep      VAWIQDSAFI RPPYFIKGIP EVKLAS---- -------
NL95 repli      VAWIQDSAFI RPPYFIKGIP EVKLAS---- -------
FI replica      VAWIQDSAFI RPPYS-TGLP EVKFARKTLV RNGKGAR
```

Appendix C
Group I and JS Nucleotides

```
Alignment: Align Group I and JS nucleotide sequences.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       GGGTGGGACC CCTTTCGGGG TCCTGCTCAA CTTCCTGTCG AGCTAAATGC
DL1 I           .......... .......... .......... .......... ..........
DL2 I           ---------- ---------- ---------- ---------- ----------
DL13 I          ---------- ---------- ---------- ---------- ----------
DL16 I          .......... .......... .......... .......... ......-...
ST4 I           .......... .......... .......... .......... ......-...
R17 I           .......... .......... .......... .......... ......-...
J20 I           .......... .......... .......... .......... ......-...
MS2 I           .......... .......... .......... .......... ......-...
M12 I           .......... .......... .......... .......... ......-...
DL52            ---------- ---------- ---------- ---------- ----------
DL54            ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       CATTTTTAAT GTCTTTAGCG AGACGCTACC WTGGCTATCG CTGTAGGTAG
DL1 I           .......... .......... .......... A......... ..........
DL2 I           ---------- ---------- ---------. A......... ..........
DL13 I          ---------- ---------- ---------. A......... ..........
DL16 I          .......... .......... .......... A......... ..........
ST4 I           .......... .......... .......... A......... ..........
R17 I           .......... .......... .......... A......... ..........
J20 I           .......... .......... .......... A......... ..........
MS2 I           .......... .......... .......... A......... ..........
M12 I           .......... .......... .......... T......... ..........
DL52            ---------- ---------- ---------- ---------- ----------
DL54            ---------- ---------- ---------- ---------- ----------

                                                    ORF1
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       CCGGAATTCC ATTCCTAGGA GGTTTGACYY RTGCGAGCTT TYAGTRYYCT
DL1 I           .......... .......... ........TC A......... .C...GTC..
DL2 I           .......... .......... ........TC A......... .T...GTC..
DL13 I          .......... .......... ........TC A......... .T...GTC..
DL16 I          .......... .......... ........TC A......... .T...GTC..
ST4 I           .......... .......... ........CT G......... .T...ACC..
R17 I           .......... .......... ........CT A......... .T...GCC..
J20 I           .......... .......... ........TC A......... .T...GTC..
MS2 I           .......... .......... ........CT G......... .T...ACC..
M12 I           .......... .......... ........TC A......... .T...GTT..
DL52            ---------- ---------- -------.TC A......... .T...GTC..
DL54            ---------- ---------- -------..TC A......... .T...GTC..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
Consensus       TGATMRGGAR WVYRARACCT TYGTSCCBHB CRTYCGCRYY TAYGCKRACG
DL1 I           ....AA...G TCTG.G.... .T..C..GCT .G.C...ACC ..T..TG...
DL2 I           ....CA...G TCCG.G.... .C..C..TAG .G.C...GTT ..T..GG...
DL13 I          ....CA...G TCCG.G.... .C..C..TAG .G.C...GTT ..T..GG...
DL16 I          ....CA...G TCCG.G.... .C..C..TAG .G.C...GTT ..T..GG...
ST4 I           ....AG...G AACG.G.... .C..C..CTC .G.T...GTT ..C..GG...
R17 I           ....AA...G AGCA.G.... .C..C..TTC .A.T...GTT ..C..GA...
J20 I           ....AG...A TCCG.G.... .C..C..CTC .G.T...GTT ..T..TG...
MS2 I           ....AG...G AACG.G.... .C..C..CTC .G.T...GTT ..C..GG...
M12 I           ....CA...G AACG.A.... .C..G..CTC .G.T...GTT ..T..GG...
DL52            ....CA...G TCCG.G.... .C..C..TAG .G.C...GTT ..T..GG...
DL54            ....CA...G TCCG.G.... .C..C..TAG .G.C...GTT ..T..GG...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
Consensus       GBSAGRYYGA RGATAACTCR TTYTCYYTVA AATAYCGYTC GAACTGGACY
DL1 I           .CG..GTT.. A........G ..C..CT.G. ....C..T.. .........C
DL2 I           .GC..GTC.. A........G ..T..CC.A. ....C..T.. .........T
DL13 I          .GC..GTC.. A........G ..T..CC.A. ....C..C.. .........T
DL16 I          .GC..GTC.. A........G ..T..CC.A. ....C..C.. .........T
ST4 I           .TG..ACT.. A........A ..C..TT.A. ....T..T.. .........T
R17 I           .TG..ACC.. A........A ..C..TT.A. ....T..C.. .........T
J20 I           .GG..GTC.. G........A ..T..CT.A. ....C..T.. .........T
MS2 I           .TG..ACT.. A........A ..C..TT.A. ....T..T.. .........T
M12 I           .CG..ACT.. A........G ..T..CC.C. ....T..C.. .........C
DL52            .GC..GTC.. A........G ..T..CC.A. ....C..C.. .........T
DL54            .GC..GTC.. A........G ..T..CC.A. ....C..C.. .........T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus       CCBGGYCGWT TTAAYTCGAC YGGGGDCYARA ACGRADCAGT GGCACTAYCC
DL1 I           ..T..T..T. ....C..... T...G.T.G. ...A.A.... .......T..
DL2 I           ..C..T..A. ....T..... C...T.T.G. ...A.A.... .......T..
DL13 I          ..C..T..A. ....T..... C...T.T.G. ...G.A.... .......T..
DL16 I          ..C..T..A. ....T..... C...T.T.G. ...G.A.... .......T..
ST4 I           ..C..T..T. ....C..... T...G.C.A. ...A.A.... .......C..
R17 I           ..C..T..T. ....C..... T...G.C.G. ...A.G.... .......T..
J20 I           ..T..C..A. ....C..... T...A.C.G. ...A.T.... .......T..
MS2 I           ..C..T..T. ....C..... T...G.C.A. ...A.A.... .......C..
M12 I           ..G..T..A. ....C..... T...G.C.G. ...A.A.... .......T..
DL52            ..C..T..A. ....T..... C...T.T.G. ...G.A.... .......T..
DL54            ..C..T..A. ....T..... C...T.T.G. ...G.A.... .......T..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       STCYYCKTAY TCDMGRGGDG CGYTNAGYGT YACDKCGRTV GATCAAGGTK
DL1 I           G..CT.T..T ..TA.A..T. ..C.T..C.. C..AT..G.G .........G
DL2 I           G..CC.T..T ..TA.G..T. ..C.C..T.. C..AT..G.A .........G
DL13 I          G..CC.T..T ..TA.G..T. ..C.C..T.. C..AT..G.A .........G
DL16 I          G..CC.T..T ..TA.G..T. ..C.C..T.. C..AT..G.A .........G
ST4 I           C..TC.G..T ..AC.G..G. ..T.A..T.. C..AT..A.A .........G
R17 I           C..CC.G..T ..GC.G..G. ..T.A..T.. C..GT..A.A .........G
J20 I           G..CC.T..C ..TA.G..A. ..C.T..T.. C..TT..G.C .........T
MS2 I           C..TC.G..T ..AC.G..G. ..T.A..T.. C..AT..A.A .........G
M12 I           C..CC.T..C ..TC.G..A. ..T.G..T.. T..TG..A.A .........G
DL52            G..CC.T..T ..TA.G..T. ..C.C..T.. C..AT..G.A .........G
DL54            G..CC.T..T ..TA.G..T. ..C.C..T.. C..AT..G.A .........G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       CYTAYAAGCG MWSTGGGTCR TCGTGGGGTC GYCCGTACGA GGAGAAARCY
DL1 I           .T..T..... CTC......A .......... .C........ .......G.C
DL2 I           .C..C..... CTC......A .......... .C........ .......G.C
DL13 I          .C..C..... CTC......A .......... .C........ .......G.C
DL16 I          .C..C..... CTC......A .......... .C........ .......G.C
ST4 I           .C..C..... AAG......A .......... .C........ .......G.C
R17 I           .C..T..... CAG......A .......... .C........ .......G.C
J20 I           .C..C..... CTC......A .......... .T........ .......G.T
MS2 I           .C..C..... AAG......A .......... .C........ .......G.C
M12 I           .C..T..... AAG......G .......... .C........ .......A.C
DL52            .C..C..... CTC......A .......... .C........ .......G.C
DL54            .C..C..... CTC......A .......... .C........ .......G.C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       GGTTWYGGYT TCTCNCTCGA CGCACGYTCC TGCTAYAGCC TCTTCCCTGT
DL1 I           ....AT..C. ....A..... ......C... .....C.... ..........
DL2 I           ....TT..C. ....A..... ......T... .....T.... ..........
DL13 I          ....TT..C. ....A..... ......T... .....T.... ..........
DL16 I          ....TT..C. ....A..... ......T... .....T.... ..........
ST4 I           ....TC..C. ....C..... ......C... .....C.... ..........
R17 I           ....TT..C. ....T..... ......C... .....C.... ..........
J20 I           ....TT..C. ....G..... ......T... .....C.... ..........
MS2 I           ....TC..C. ....C..... ......C... .....C.... ..........
M12 I           ....TT..T. ....A..... ......C... .....C.... ..........
DL52            ....TT..C. ....A..... ......T... .....T.... ..........
DL54            ....TT..C. ....A..... ......T... .....T.... ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       HAGYCARAAY HTGACTTACA TYGAAGTGCC GCAGAACGTT GCGAAYCGGG
DL1 I           T..T..G..C C......... .T........ .......... .....T....
DL2 I           T..T..G..T C......... .C........ .......... .....C....
DL13 I          T..T..G..T C......... .C........ .......... .....C....
DL16 I          T..T..G..T C......... .C........ .......... .....C....
ST4 I           A..C..G..C T......... .C........ .......... .....C....
R17 I           A..C..G..C T......... .C........ .......... .....T....
J20 I           C..T..A..T A......... .C........ .......... .....T....
MS2 I           A..C..A..C T......... .C........ .......... .....C....
M12 I           A..C..G..C C......... .C........ .......... .....C....
DL52            T..T..G..T C......... .C........ .......... .....C....
DL54            T..T..G..T C......... .C........ .......... .....C....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus       CKWCGACCGA AGTCCTGCAR AAGGTYACYC ARGGNAAYTT YAACCTTGGB
DL1 I           .GT....... .........G .....C..C. .A..C..T.. C........T
DL2 I           .GT....... .........G .....C..T. .A..A..T.. C........T
DL13 I          .GT....... .........G .....C..T. .A..A..T.. C........T
DL16 I          .GT....... .........G .....C..T. .A..A..T.. C........T
ST4 I           .GT....... .........A .....C..C. .G..T..T.. T........T
R17 I           .GT....... .........A .....T..C. .A..T..T.. C........T
J20 I           .GT....... .........G .....T..C. .G..T..T.. T........C
MS2 I           .GT....... .........A .....C..C. .G..T..T.. T........T
M12 I           .TA....... .........G .....T..C. .A..G..C.. C........G
DL52            .GT....... .........G .....C..T. .A..A..T.. C........T
DL54            .GT....... .........G .....C..T. .A..A..T.. C........T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       560        570        580        590        600
Consensus       GTNGCYYTWG CAGAGGCVAG RTCGACAGCC TCACAACTCG CGACGCAAAC
DL1 I           ..A..CT.A. .......G.. G......... .......... ..........
DL2 I           ..G..CC.A. .......C.. G......... .......... ..........
DL13 I          ..G..CC.A. .......C.. G......... .......... ..........
DL16 I          ..G..CC.A. .......C.. G......... .......... ..........
ST4 I           ..T..TT.A. .......C.. G......... .......... ..........
R17 I           ..C..CC.A. .......C.. A......... .......... ..........
J20 I           ..A..TT.A. .......C.. G......... .......... ..........
MS2 I           ..T..TT.A. .......C.. G......... .......... ..........
M12 I           ..G..CC.T. .......A.. G......... .......... ..........
DL52            ..G..CC.A. .......C.. G......... .......... ..........
DL54            ..G..CC.A. .......C.. G......... .......... ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       610        620        630        640        650
Consensus       CATTGCGCTC GTGAAGGCGT ACACTGCCGC TCGTCGCGGY AAYTGGCGCC
DL1 I           .......... .......... .......... .........C ..C.......
DL2 I           .......... .......... .......... .........C ..T.......
DL13 I          .......... .......... .......... .........C ..T.......
DL16 I          .......... .......... .......... .........C ..T.......
ST4 I           .......... .......... .......... .........T ..T.......
R17 I           .......... .......... .......... .........T ..T.......
J20 I           .......... .......... .......... .........T ..T.......
MS2 I           .......... .......... .......... .........T ..T.......
M12 I           .......... .......... .......... .........T ..T.......
DL52            .......... .......... .......... .........C ..T.......
DL54            .......... .......... .......... .........C ..T.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       660        670        680        690        700
Consensus       AGVCGSTCCG CTAYCTYGCC CTWAACGAAG AYCGRAARTT YCGRTCRAAA
DL1 I           ..A..C.... ...T..C... ..T....... .C..A..A.. T..G..A...
DL2 I           ..A..C.... ...T..T... ..A....... .T..G..G.. C..A..G...
DL13 I          ..A..C.... ...T..T... ..A....... .T..G..G.. C..A..G...
DL16 I          ..A..C.... ...T..T... ..A....... .T..G..G.. C..A..G...
ST4 I           ..G..C.... ...C..T... ..A....... .T..A..G.. T..A..A...
R17 I           ..G..C.... ...C..C... ..A....... .T..A..A.. T..A..A...
J20 I           ..G..C.... ...T..T... ..A....... .T..G..A.. C..G..A...
MS2 I           ..G..C.... ...C..T... ..A....... .T..A..G.. T..A..A...
M12 I           ..C..G.... ...C..C... ..A....... .T..A..G.. T..A..A...
DL52            ..A..C.... ...T..T... ..A....... .T..G..G.. C..A..G...
DL54            ..A..C.... ...T..T... ..A....... .T..G..G.. C..A..G...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       710        720        730        740        750
Consensus       CACGTGGCVG GYAGRTGGTT GGAGTTGCAG TTCGGHTGGY TACCRCTHAT
DL1 I           ........G. .C..G..... .......... .....C...T ....G..C..
DL2 I           ........A. .C..G..... .......... .....C...C ....G..C..
DL13 I          ........A. .C..G..... .......... .....C...C ....G..C..
DL16 I          ........A. .C..G..... .......... .....C...C ....G..C..
ST4 I           ........C. .C..G..... .......... .....T...T ....A..A..
R17 I           ........C. .C..G..... .......... .....A...T ....A..A..
J20 I           ........A. .T..A..... .......... .....T...C ....G..T..
MS2 I           ........C. .C..G..... .......... .....T...T ....A..A..
M12 I           ........C. .C..G..... .......... .....C...T ....A..T..
DL52            ........A. .C..G..... .......... .....C...C ....G..C..
DL54            ........A. .C..G..... .......... .....C...C ....G..C..
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       760        770        780        790        800
Consensus       GAGYGATATC CARGGYGCAT AYGAGATGCT TACGAAGGTT CAYCTTCAAG
DL1 I           ...C...... ..G..C.... .T........ .......... ..C.......
DL2 I           ...C...... ..G..C.... .T........ .......... ..C.......
DL13 I          ...C...... ..G..C.... .T........ .......... ..C.......
DL16 I          ...C...... ..G..C.... .T........ .......... ..C.......
ST4 I           ...T...... ..G..T.... .T........ .......... ..C.......
R17 I           ...T...... ..A..T.... .T........ .......... ..C.......
J20 I           ...C...... ..A..T.... .T........ .......... ..T.......
MS2 I           ...T...... ..G..T.... .T........ .......... ..C.......
M12 I           ...T...... ..G..C.... .C........ .......... ..C.......
DL52            ...C...... ..G..C.... .T........ .......... ..C.......
DL54            ...C...... ..G..C.... .T........ .......... ..C.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       810        820        830        840        850
Consensus       AGTTTCTYCC TATGAGRGCC GTRCGNCARG TNGGYACTAA CRTYAAGTTA
DL1 I           .......C.. ......A... ..G..C..G. .A..C..... .G.C......
DL2 I           .......T.. ......G... ..A..G..A. .G..C..... .A.T......
DL13 I          .......T.. ......G... ..A..G..A. .G..C..... .A.T......
DL16 I          .......T.. ......G... ..A..G..A. .G..C..... .A.T......
ST4 I           .......T.. ......A... ..A..T..G. .C..T..... .A.C......
R17 I           .......T.. ......A... ..A..T..A. .T..T..... .A.T......
J20 I           .......T.. ......A... ..A..A..G. .G..T..... .G.T......
MS2 I           .......T.. ......A... ..A..T..G. .C..T..... .A.C......
M12 I           .......C.. ......A... ..A..T..G. .C..C..... .A.T......
DL52            .......T.. ......G... ..A..G..A. .G..C..... .A.T......
DL54            .......T.. ......G... ..A..G..A. .G..C..... .A.T......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       860        870        880        890        900
Consensus       RATGGCCGYY TKKCGTATCC AGCTGCAAAC TWCCARACWA CGTGCAACAT
DL1 I           G.......CC .GT....... .......... .A...A..T. ..........
DL2 I           G.......TT .GT....... .......... .A...G..T. ..........
DL13 I          G.......TT .GT....... .......... .A...G..T. ..........
DL16 I          G.......TT .GT....... .......... .A...G..T. ..........
ST4 I           A.......TC .GT....... .......... .T...G..A. ..........
R17 I           G.......CT .GG....... .......... .T...G..A. ..........
J20 I           G.......CT .GT....... .......... .A...G..T. ..........
MS2 I           G.......TC .GT....... .......... .T...G..A. ..........
M12 I           G.......CC .TT....... .......... .A...G..A. ..........
DL52            G.......TT .GT....... .......... .A...G..T. ..........
DL54            G.......TT .GT....... .......... .A...G..T. ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       910        920        930        940        950
Consensus       ATCRCGACGW ATCGTGATAT GGTTTTACAT AAACGATGCA CGWTTGGCHT
DL1 I           ...G.....T .......... .......... .......... ..T.....T.
DL2 I           ...A.....A .......... .......... .......... ..T.....C.
DL13 I          ...A.....A .......... .......... .......... ..T.....C.
DL16 I          ...A.....A .......... .......... .......... ..T.....C.
ST4 I           ...G.....T .......... .......... .......... ..T.....A.
```

```
R17 I        ...A.....T ......... ......... ......... ..T.....A.
J20 I        ...G.....A ......... ......... ......... ..T.....C.
MS2 I        ...G.....T ......... ......... ......... ..T.....A.
M12 I        ...G.....T ......... ......... ......... ..A.....C.
DL52         ...A.....A ......... ......... ......... ..T.....C.
DL54         ...A.....A ......... ......... ......... ..T.....C.


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            960        970        980        990       1000
Consensus    GGTTGTCGTC TCTRGGTATC TTGAACCCAC TAGGTATAGT GTGGGAAAAG
DL1 I        .......... ...A...... .......... .......... ..........
DL2 I        .......... ...A...... .......... .......... ..........
DL13 I       .......... ...A...... .......... .......... ..........
DL16 I       .......... ...A...... .......... .......... ..........
ST4 I        .......... ...A...... .......... .......... ..........
R17 I        .......... ...A...... .......... .......... ..........
J20 I        .......... ...A...... .......... .......... ..........
MS2 I        .......... ...A...... .......... .......... ..........
M12 I        .......... ...G...... .......... .......... ..........
DL52         .......... ...A...... .......... .......... ..........
DL54         .......... ...A...... .......... .......... ..........


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           1010       1020       1030       1040       1050
Consensus    GTGCCYTTCT CATTCGTTGT CGACTGGCTC CTDCCTGTDG GDAACATGCT
DL1 I        .....T.... .......... .......... ..T.....A. .A........
DL2 I        .....T.... .......... .......... ..T.....T. .G........
DL13 I       .....T.... .......... .......... ..T.....T. .G........
DL16 I       .....T.... .......... .......... ..T.....T. .G........
ST4 I        .....T.... .......... .......... ..A.....A. .T........
R17 I        .....T.... .......... .......... ..A.....A. .T........
J20 I        .....T.... .......... .......... ..T.....A. .A........
MS2 I        .....T.... .......... .......... ..A.....A. .T........
M12 I        .....C.... .......... .......... ..G.....G. .G........
DL52         .....T.... .......... .......... ..T.....T. .G........
DL54         .....T.... .......... .......... ..T.....T. .G........


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                           1060       1070       1080       1090       1100
Consensus    MGAGGGCCTH ACRGCYCCMG TDGGATGYTC YTACATGTCD GGRACAGTTA
DL1 I        C........A ..G..C..C. .A.....T.. T........G ..A.......
DL2 I        C........C ..A..C..C. .A.....C.. C........T ..A.......
DL13 I       C........C ..A..C..C. .A.....C.. C........T ..A.......
DL16 I       C........C ..A..C..C. .A.....C.. C........T ..A.......
ST4 I        C........T ..G..C..C. .G.....C.. C........A ..A.......
R17 I        C........T ..G..T..C. .T.....C.. C........A ..A.......
J20 I        C........T ..A..C..C. .A.....C.. C........T ..A.......
MS2 I        C........T ..G..C..C. .G.....C.. C........A ..A.......
M12 I        A........T ..G..T..A. .A.....T.. T........A ..G.......
DL52         C........C ..A..C..C. .A.....C.. C........T ..A.......
DL54         C........C ..A..C..C. .A.....C.. C........T ..A.......
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1110       1120       1130       1140       1150
Consensus       CTGACGTAAT AACGGGTGAG TCCATCATAA GCGTYGACGC TCCCTAYGGG
DL1 I           .......... .......... .......... ....T..... ......T...
DL2 I           .......... .......... .......... ....C..... ......T...
DL13 I          .......... .......... .......... ....C..... ......T...
DL16 I          .......... .......... .......... ....C..... ......T...
ST4 I           .......... .......... .......... ....T..... ......C...
R17 I           .......... .......... .......... ....T..... ......T...
J20 I           .......... .......... .......... ....T..... ......T...
MS2 I           .......... .......... .......... ....T..... ......C...
M12 I           .......... .......... .......... ....T..... ......C...
DL52            .......... .......... .......... ....C..... ......T...
DL54            .......... .......... .......... ....C..... ......T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1160       1170       1180       1190       1200
Consensus       TGGACDGTGG AKAGACAGGG CACTGCTAAG GCCCADRTYT CRGCCATGCA
DL1 I           .....A.... .G........ .......... .....GA.C. .A........
DL2 I           .....G.... .T........ .......... .....AG.C. .A........
DL13 I          .....G.... .T........ .......... .....AG.C. .A........
DL16 I          .....G.... .T........ .......... .....AG.C. .A........
ST4 I           .....T.... .G........ .......... .....AA.C. .A........
R17 I           .....T.... .G........ .......... .....TG.T. .A........
J20 I           .....T.... .G........ .......... .....GA.C. .G........
MS2 I           .....T.... .G........ .......... .....AA.C. .A........
M12 I           .....T.... .G........ .......... .....AG.C. .A........
DL52            .....G.... .T........ .......... .....AG.C. .A........
DL54            .....G.... .T........ .......... .....AG.C. .A........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1210       1220       1230       1240       1250
Consensus       TCGAGGGGTR CARTCCGTAT GSCCAACWAC TGGCGYRTAC GTRAARTCWC
DL1 I           .........A ..G....... .G.....A.. .....TA... ..A..G..A.
DL2 I           .........G ..A....... .G.....A.. .....TA... ..G..G..A.
DL13 I          .........G ..A....... .G.....A.. .....TA... ..G..G..A.
DL16 I          .........G ..A....... .G.....A.. .....TA... ..G..G..A.
ST4 I           .........A ..A....... .G.....A.. .....CG... ..A..G..T.
R17 I           .........A ..A....... .G.....A.. .....CA... ..A..G..T.
J20 I           .........G ..A....... .G.....A.. .....TA... ..G..A..A.
MS2 I           .........A ..A....... .G.....A.. .....CG... ..A..G..T.
M12 I           .........A ..A....... .C.....T.. .....TA... ..A..G..T.
DL52            .........G ..A....... .G.....A.. .....TA... ..G..G..A.
DL54            .........G ..A....... .G.....A.. .....TA... ..G..G..A.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1260       1270       1280       1290       1300
Consensus       CYTTYTCGAT KGTCCAYACY TTAGAYGCGT TRGCATTAAT CAGGCAACGG
DL1 I           .T..C..... G.....C..T .....T.... .G........ ..........
DL2 I           .T..C..... G.....T..C .....C.... .G........ ..........
DL13 I          .T..C..... G.....T..C .....C.... .G........ ..........
DL16 I          .T..C..... G.....T..C .....C.... .G........ ..........
ST4 I           .T..C..... G.....T..C .....T.... .A........ ..........
```

```
R17 I        .T..C..... G.....T..C .....T.... .A........ ..........
J20 I        .T..C..... T.....T..C .....T.... .A........ ..........
MS2 I        .T..C..... G.....T..C .....T.... .A........ ..........
M12 I        .C..T..... G.....T..C .....T.... .G........ ..........
DL52         .T..C..... G.....T..C .....C.... .G........ ..........
DL54         .T..C..... G.....T..C .....C.... .G........ ..........

                      STOP 1                          ORF2
             ....|....| ....|....| ....|....| ....|....| ....|....|
                     1310       1320       1330       1340       1350
Consensus    CTCTYWARAT AGAGCCCTCA ACCGGAGTTY GAAGCATGGC TTCTAACTTT
DL1 I        ....CA.G.. .......... .........C .......... ..........
DL2 I        ....TA.G.. .......... .........C .......... ..........
DL13 I       ....TA.G.. .......... .........C .......... ..........
DL16 I       ....TA.G.. .......... .........C .......... ..........
ST4 I        ....CT.G.. .......... .........T .......... ..........
R17 I        ....CT.A.. .......... .........T .......... ..........
J20 I        ....CA.G.. .......... .........T .......... ..........
MS2 I        ....CT.G.. .......... .........T .......... ..........
M12 I        ....CT.A.. .......... .........T .......... ..........
DL52         ....TA.G.. .......... .........C .......... ..........
DL54         ....TA.G.. .......... .........C .......... ..........


             ....|....| ....|....| ....|....| ....|....| ....|....|
                     1360       1370       1380       1390       1400
Consensus    ACTCAGTTYG TTCTCGTCGA CAATGGCGGA ACYGGYGACG TGACTGTCGC
DL1 I        ........C. .......... .......... ..T..T.... ..........
DL2 I        ........C. .......... .......... ..T..C.... ..........
DL13 I       ........C. .......... .......... ..T..C.... ..........
DL16 I       ........C. .......... .......... ..T..C.... ..........
ST4 I        ........C. .......... .......... ..T..C.... ..........
R17 I        ........T. .......... .......... ..T..C.... ..........
J20 I        ........T. .......... .......... ..C..C.... ..........
MS2 I        ........C. .......... .......... ..T..C.... ..........
M12 I        ........T. .......... .......... ..T..C.... ........N.
DL52         ........C. .......... .......... ..T..C.... ..........
DL54         ........C. .......... .......... ..T..C.... ..........


             ....|....| ....|....| ....|....| ....|....| ....|....|
                     1410       1420       1430       1440       1450
Consensus    BCCAAGCAAC TTCGCTAACG GGGTCGCTGA ATGGATYAGC TCYAACTCVC
DL1 I        C......... .......... .......... ......T... ..T.....G.
DL2 I        T......... .......... .......... ......T... ..T.....A.
DL13 I       T......... .......... .......... ......T... ..T.....A.
DL16 I       T......... .......... .......... ......T... ..T.....A.
ST4 I        C......... .......... .......... ......C... ..T.....G.
R17 I        T......... .......... .......... ......C... ..T.....G.
J20 I        T......... .......... .......... ......C... ..T.....A.
MS2 I        C......... .......... .......... ......C... ..T.....G.
M12 I        G......... .......... .......... ......C... ..C.....C.
DL52         T......... .......... .......... ......T... ..T.....A.
DL54         T......... .......... .......... ......T... ..T.....A.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1460       1470       1480       1490       1500
Consensus       GYTCWCARGC TTACAAAGTA ACYTGTAGYG TKCGTCAGAG CTCTGCGCAG
DL1 I           .C..T..G.. .......... ..C.....C. .T........ ..........
DL2 I           .T..T..G.. .......... ..C.....C. .T........ ..........
DL13 I          .T..T..G.. .......... ..C.....C. .T........ ..........
DL16 I          .T..T..G.. .......... ..C.....C. .T........ ..........
ST4 I           .T..A..G.. .......... ..C.....C. .T........ ..........
R17 I           .C..A..G.. .......... ..C.....C. .T........ ..........
J20 I           .C..T..G.. .......... ..C.....T. .T........ ..........
MS2 I           .T..A..G.. .......... ..C.....C. .T........ ..........
M12 I           .C..A..A.. .......... ..T.....C. .G........ ..........
DL52            .T..T..G.. .......... ..C.....C. .T........ ..........
DL54            .T..T..G.. .......... ..C.....C. .T........ ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1510       1520       1530       1540       1550
Consensus       AAYCGCAART ACACYATYAA RGTYGARGTR CCDAARGTGG CWACYCARAC
DL1 I           ..T.....G. ....T..C.. G..C..A..G ..G..A.... .T..C..G..
DL2 I           ..C.....G. ....C..T.. G..T..G..A ..A..A.... .T..C..A..
DL13 I          ..C.....G. ....C..T.. G..T..G..A ..A..A.... .T..C..A..
DL16 I          ..C.....G. ....C..T.. G..T..G..A ..A..A.... .T..C..A..
ST4 I           ..T.....A. ....C..C.. A..C..G..G ..T..A.... .A..C..G..
R17 I           ..T.....A. ....C..T.. A..C..G..G ..T..G.... .A..T..G..
J20 I           ..C.....G. ....C..C.. A..C..A..G ..A..A.... .T..T..G..
MS2 I           ..T.....A. ....C..C.. A..C..G..G ..T..A.... .A..C..G..
M12 I           ..C.....A. ....C..C.. G..C..G..G ..G..A.... .A..C..A..
DL52            ..C.....G. ....C..T.. G..T..G..A ..A..A.... .T..C..A..
DL54            ..C.....G. ....C..T.. G..T..G..A ..A..A.... .T..C..A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       YGTYGGYGGY GTASAGCTTC CTGTAGCCGC RTGGCGTTCG TAYYTRAATA
DL1 I           C..T..T..C ...C...... .......... G......... ..TC.G....
DL2 I           C..C..T..T ...C...... .......... A......... ..CT.G....
DL13 I          C..C..T..T ...C...... .......... A......... ..CT.G....
DL16 I          C..C..T..T ...C...... .......... A......... ..CT.G....
ST4 I           T..T..T..T ...G...... .......... A......... ..CT.A....
R17 I           T..T..T..T ...G...... .......... A......... ..CT.A....
J20 I           C..C..T..C ...C...... .......... A......... ..CT.A....
MS2 I           T..T..T..T ...G...... .......... A......... ..CT.A....
M12 I           C..T..C..T ...G...... .......... A......... ..CC.G....
DL52            C..C..T..T ...C...... .......... A......... ..CT.G....
DL54            C..C..T..T ...C...... .......... A......... ..CT.G....
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        1610       1620       1630       1640       1650
Consensus        TGGAAYTRAC YATTCCDATY TTCGCDACGA AYKMCGAYTG CGMGCTWATT
DL1 I            .....T.G.. T.....A..C .....A.... .CGA...C.. ..C...A...
DL2 I            .....T.A.. T.....T..T .....A.... .CGA...T.. ..C...T...
DL13 I           .....T.A.. T.....T..T .....A.... .CGA...T.. ..C...T...
DL16 I           .....T.A.. T.....T..C .....A.... .CGA...T.. ..C...T...
ST4 I            .....C.A.. C.....A..T .....T.... .TTC...C.. ..A...T...
R17 I            .....T.A.. T.....A..T .....T.... .CTC...T.. ..A...T...
J20 I            .....T.A.. T.....A..T .....G.... .CGA...T.. ..C...T...
MS2 I            .....C.A.. C.....A..T .....T.... .TTC...C.. ..A...T...
M12 I            .....T.A.. T.....G..T .....T.... .CTC...T.. ..C...T...
DL52             .....T.A.. T.....T..T .....A.... .CGA...T.. ..C...T...
DL54             .....T.A.. T.....T..T .....A.... .CGA...T.. ..C...T...
                                                ORF3
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        1660       1670       1680       1690       1700
Consensus        GTYAARGCRA TGCAAGGTCT CCTRAAAGAT GGAAACCCRA TYCCYTCRGC
DL1 I            ..T..G..G. .......... ...A...... ........G. .C..C..A..
DL2 I            ..T..G..G. .......... ...A...... ........G. .C..C..A..
DL13 I           ..T..G..G. .......... ...A...... ........G. .C..C..A..
DL16 I           ..T..G..G. .......... ...A...... ........G. .C..C..A..
ST4 I            ..T..G..A. .......... ...A...... ........G. .T..C..A..
R17 I            ..T..G..A. .......... ...A...... ........G. .T..C..A..
J20 I            ..T..G..G. .......... ...A...... ........A. .C..C..A..
MS2 I            ..T..G..A. .......... ...A...... ........G. .T..C..A..
M12 I            ..C..A..A. .......... ...G...... ........G. .T..T..G..
DL52             ..T..G..G. .......... ...A...... ........G. .C..C..A..
DL54             ..T..G..G. .......... ...A...... ........G. .C..C..A..


                                              STOP 2
               ....|....| ....|....| ....|....| ....|....| ....|....|
                        1710       1720       1730       1740       1750
Consensus        AATCGCAGCM AACTCCGGMA TCTAYTAATA GAYDYCGKCC ATTCMAACAT
DL1 I            .........A ........A. ....C..... ..TTT..G.. ....A.....
DL2 I            .........A ........A. ....C..... ..TTC..T.. ....C.....
DL13 I           .........A ........A. ....C..... ..TTC..T.. ....C.....
DL16 I           .........A ........A. ....C..... ..TTC..T.. ....C.....
ST4 I            .........A ........C. ....C..... ..CGC..G.. ....A.....
R17 I            .........A ........C. ....C..... ..TGC..G.. ....A.....
J20 I            .........C ........A. ....T..... ..TTC..G.. ....C.....
MS2 I            .........A ........C. ....C..... ..CGC..G.. ....A.....
M12 I            .........A ........C. ....C..... ..TAC..G.. ....A.....
DL52             .........A ........A. ....C..... ..TTC..T.. ....C.....
DL54             .........A ........A. ....C..... ..TTC..T.. ....C.....

                            ORF4
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        1760       1770       1780       1790       1800
Consensus        GAGGATTACC CATGTCGAAG ACAACAAAGA AGTTCAACTC TTTATGTATT
DL1 I            .......... .......... .......... .......... ..........
DL2 I            .......... .......... .......... .......... ..........
DL13 I           .......... .......... .......... .......... ..........
DL16 I           .......... .......... .......... .......... ..........
ST4 I            .......... .......... .......... .......... ..........
```

```
R17 I          .......... .......... .......... .......... ..........
J20 I          .......... .......... .......... .......... ..........
MS2 I          .......... .......... .......... .......... ..........
M12 I          .......... .........R .......... .......... ..........
DL52           .......... .......... .......... .......... ..........
DL54           .......... .......... .......... .......... ..........


                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1810       1820       1830       1840       1850
Consensus      GATCTTCCTY GCGATCTTTC TCTCGAAATT TACCAATCAA TTGCTTCTGT
DL1 I          .........C .......... .......... .......... ..........
DL2 I          .........C .......... .......... .......... ..........
DL13 I         .........C .......... .......... .......... ..........
DL16 I         .........T .......... .......... .......... ..........
ST4 I          .........C .......... .......... .......... ..........
R17 I          .........C .......... .......... .......... ..........
J20 I          .........C .......... .......... .......... ..........
MS2 I          .........C .......... .......... .......... ..........
M12 I          .........C .......... .......... .......... ..........
DL52           .........C .......... .......... .......... ..........
DL54           .........C .......... .......... .......... ..........


                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1860       1870       1880       1890       1900
Consensus      CGCTACTGGA WGCGGTRATC CGCACAGTRR MGACTTTWCR GCAATTGCTT
DL1 I          .......... A.....G... ........AG A......A.A ..........
DL2 I          .......... A.....G... ........AG A......A.A ..........
DL13 I         .......... A.....G... ........AG A......A.A ..........
DL16 I         .......... A.....G... ........AG A......A.A ..........
ST4 I          .......... A.....G... ........GA C......A.A ..........
R17 I          .......... A.....A... ........GA C......A.A ..........
J20 I          .......... A.....G... ........AG A......A.G ..........
MS2 I          .......... A.....G... ........GA C......A.A ..........
M12 I          .......... T.....A... ........GA C......T.A ..........
DL52           .......... A.....G... ........AG A......A.A ..........
DL54           .......... A.....G... ........AG A......A.A ..........

                             STOP 3
                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1910       1920       1930       1940       1950
Consensus      ACYTAAGRGA YGARTTGCTH ACWAAGCAYC CVWCNTTRGG HWMYGGTAAT
DL1 I          ..C....G.. C..G.....A ..T.....T. .CT.C..A.. CAAT......
DL2 I          ..C....A.. C..A.....A ..T.....T. .GT.C..A.. AAAT......
DL13 I         ..C....A.. C..A.....A ..T.....T. .GT.C..A.. AAAT......
DL16 I         ..C....A.. C..A.....A ..T.....T. .GT.C..A.. AAAT......
ST4 I          ..T....G.. C..A.....C ..A.....T. .GA.C..A.. TTCT......
R17 I          ..C....G.. C..A.....T ..A.....T. .GA.T..A.. CTCT......
J20 I          ..C....A.. C..G.....C ..A.....T. .GT.T..A.. AAAT......
MS2 I          ..T....G.. C..A.....C ..A.....T. .GA.C..A.. TTCT......
M12 I          ..C....G.. C..G.....T ..T.....C. .GA.G..A.. TTCC......
DL52           ..C....A.. C..A.....A ..T.....T. .GT.C..A.. AAAT......
DL54           ..C....A.. T..G.....T ..T.....T. .AT.A..G.. CAAT......
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1960       1970       1980       1990       2000
Consensus       GACGAGGCGA CCCGYCGNRC YYTAGCTATY GCTAAGCTNC GGGAGGCGAA
DL1 I           .......... ....C..AG. TT.......T ........T. ..........
DL2 I           .......... ....C..GG. TC.......C ........C. ..........
DL13 I          .......... ....C..GG. TC.......C ........C. ..........
DL16 I          .......... ....C..GG. TC.......C ........C. ..........
ST4 I           .......... ....T..TA. CT.......C ........A. ..........
R17 I           .......... ....T..CA. CT.......T ........G. ..........
J20 I           .......... ....C..GG. TT.......C ........C. ..........
MS2 I           .......... ....T..TA. CT.......C ........A. ..........
M12 I           .......... ....C..TG. TT.......T ........A. ..........
DL52            .......... ....C..GG. TC.......C ........C. ..........
DL54            .......... ....C..GG. TC.......T ........C. ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2010       2020       2030       2040       2050
Consensus       TGRDSRDYGY GGYCAGATHA AYAGRGARGG TTTCTTACAY GAYAAATCCT
DL1 I           ..AACGAT.T ..C.....C. .T..A..G.. .........T ..C.......
DL2 I           ..AGCGGT.C ..C.....T. .T..G..A.. .........C ..T.......
DL13 I          ..AGCGGT.C ..C.....T. .T..G..A.. .........C ..T.......
DL16 I          ..AGCGGT.C ..C.....T. .T..G..A.. .........C ..T.......
ST4 I           ..ATCGGT.C ..T.....A. .T..A..A.. .........T ..C.......
R17 I           ..ATCGGT.C ..C.....A. .T..A..A.. .........T ..C.......
J20 I           ..AACGAT.T ..C.....T. .C..A..A.. .........T ..C.......
MS2 I           ..GTGATC.C ..T.....A. .T..A..A.. .........T ..C.......
M12 I           ..ATCGAT.T ..C.....T. .C..A..A.. .........T ..C.......
DL52            ..AGCGGT.C ..C.....T. .T..G..A.. .........C ..T.......
DL54            ..AGCGAT.T ..C.....T. .T..A..A.. .........T ..C.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2060       2070       2080       2090       2100
Consensus       TRTCRTGGGA TCCGGATGTT TTACAAACCA GCATCCGTAG CCTWATHGGC
DL1 I           .G..A..... .......... .......... .......... ...T..T...
DL2 I           .A..G..... .......... .......... .......... ...T..T...
DL13 I          .A..G..... .......... .......... .......... ...T..T...
DL16 I          .A..G..... .......... .......... .......... ...T..T...
ST4 I           .G..A..... .......... .......... .......... ...T..T...
R17 I           .G..A..... .......... .......... .......... ...T..C...
J20 I           .G..G..... .......... .......... .......... ...T..T...
MS2 I           .G..A..... .......... .......... .......... ...T..T...
M12 I           .G..G..... .......... .......... .......... ...A..A...
DL52            .A..G..... .......... .......... .......... ...T..T...
DL54            .G..A..... .......... .......... .......... ...T..T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        2110       2120       2130       2140       2150
Consensus       AAYCTYCTCT CTGGYTAYCR HTCGTCGTTG TTTGGGCAAT GCACGTTYTC
DL1 I           ..C..T.... ....C..T.G T......... .......... .......C..
DL2 I           ..C..T.... ....C..C.G C......... .......... .......C..
DL13 I          ..C..T.... ....C..C.G C......... .......... .......C..
DL16 I          ..C..T.... ....C..C.G C......... .......... .......C..
ST4 I           ..C..C.... ....C..C.G A......... .......... .......C..
```

```
R17 I        ..T..T.... ....C..C.A A......... .......... .......C..
J20 I        ..C..T.... ....C..T.G T......... .......... .......T..
MS2 I        ..C..C.... ....C..C.G A......... .......... .......C..
M12 I        ..C..C.... ....T..C.G A......... .......... .......C..
DL52         ..C..T.... ....C..C.G C......... .......... .......C..
DL54         ..C..C.... ....C..T.G T......... .......... .......C..


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            2160       2170       2180       2190       2200
Consensus    CAACGGTGCY YCDATGGGGC ACAAGTTGCA GGATGCAGCG CCTTACAAGA
DL1 I        .........C T.G....... .......... .......... ..........
DL2 I        .........C T.A....... .......... .......... ..........
DL13 I       .........C T.A....... .......... .......... ..........
DL16 I       .........C T.A....... .......... .......... ..........
ST4 I        .........C T.T....... .......... .......... ..........
R17 I        .........C T.T....... .......... .......... ..........
J20 I        .........C T.T....... .......... .......... ..........
MS2 I        .........T C.T....... .......... .......... ..........
M12 I        .........C T.G....... .......... .......... ..........
DL52         .........C T.A....... .......... .......... ..........
DL54         .........C T.G....... .......... .......... ..........


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            2210       2220       2230       2240       2250
Consensus    AGTTCGCTGA ACAAGCAACC GTTACCCCCC GCGCTCTRAG AGCGGCNYTR
DL1 I        .......... .......... .......... .......A.. ......CC.A
DL2 I        .......... .......... .......... .......G.. ......AC.A
DL13 I       .......... .......... .......... .......G.. ......AC.A
DL16 I       .......... .......... .......... .......G.. ......AC.A
ST4 I        .......... .......... .......... .......G.. ......TC.A
R17 I        .......... .......... .......... .......G.. ......TC.A
J20 I        .......... .......... .......... .......G.. ......GT.A
MS2 I        .......... .......... .......... .......G.. ......TC.A
M12 I        .......... .......... .......... .......A.. ......CC.G
DL52         .......... .......... .......... .......G.. ......AC.A
DL54         .......... .......... .......... .......G.. ......TC.A


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            2260       2270       2280       2290       2300
Consensus    YTGGTCMGAG ACCARTGTGY GCCGTGGATY AGACACGCGG TCCRCTAYAR
DL1 I        C.....C... ....G....T .........T .......... ...G...C.A
DL2 I        C.....C... ....G....T .........C .......... ...G...C.A
DL13 I       C.....C... ....G....T .........C .......... ...G...C.A
DL16 I       C.....C... ....G....T .........C .......... ...G...C.A
ST4 I        T.....C... ....A....C .........C .......... ...G...T.A
R17 I        T.....C... ....A....C .........C .......... ...A...T.A
J20 I        C.....A... ....G....C .........C .......... ...G...T.A
MS2 I        T.....C... ....A....C .........C .......... ...G...T.A
M12 I        C.....A... ....G....C .........T .......... ...G...C.A
DL52         C.....C... ....G....T .........C .......... ...G...C.A
DL54         C.....C... ....G....C .........C .......... ...A...C.G
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2310       2320       2330       2340       2350
Consensus       YGARTCATAT RAATTTAGRC TCGTYGTAGG GAACGGAGTG TTYACAGTTC
DL1 I           T..G...... G.......G. ....T..... .......... ..T.......
DL2 I           C..G...... G.......G. ....C..... .......... ..T.......
DL13 I          C..G...... G.......G. ....C..... .......... ..T.......
DL16 I          C..G...... G.......G. ....C..... .......... ..T.......
ST4 I           C..G...... G.......G. ....T..... .......... ..T.......
R17 I           C..G...... G.......G. ....T..... .......... ..T.......
J20 I           C..A...... A.......A. ....T..... .......... ..T.......
MS2 I           C..G...... G.......G. ....T..... .......... ..T.......
M12 I           C..G...... G.......G. ....T..... .......... ..C.......
DL52            C..G...... G.......G. ....C..... .......... ..T.......
DL54            C..G...... G.......G. ....C..... .......... ..C.......


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2360       2370       2380       2390       2400
Consensus       CGAAGAATAA TAAAATAGAT CGGGCTGCYT GYAARGAGCC HGATRYGAAY
DL1 I           .......... .......... ........C. .T..G..... C...AT...T
DL2 I           .......... .......... ........C. .T..G..... C...AT...T
DL13 I          .......... .......... ........C. .T..G..... C...AT...T
DL16 I          .......... .......... ........C. .T..G..... C...AT...T
ST4 I           .......... .......... ........C. .T..G..... T...AT...T
R17 I           .......... .......... ........C. .T..G..... T...AT...T
J20 I           .......... .......... ........C. .T..G..... C...AT...T
MS2 I           .......... .......... ........C. .T..G..... T...AT...T
M12 I           .......... .......... ........T. .C..A..... C...AT...T
DL52            .......... .......... ........C. .C..G..... T...GC...C
DL54            .......... .......... ........C. .T..G..... A...GC...C


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2410       2420       2430       2440       2450
Consensus       ATGTACCTYC AGAAAGGRGT HGGHGVYTTY ATHMGVCGYC GBCTCMRHTC
DL1 I           ........T. .......G.. T..C.CC..C ..TA.A..C. .C...AAA..
DL2 I           ........C. .......A.. T..C.CC..C ..AA.G..T. .C...AAA..
DL13 I          ........C. .......A.. T..C.CC..C ..AA.G..T. .C...AAA..
DL16 I          ........C. .......A.. T..C.CC..C ..AA.G..T. .C...AAA..
ST4 I           ........C. .......G.. C..T.CC..T ..CA.A..C. .G...AAA..
R17 I           ........C. .......G.. C..C.CT..T ..TA.A..C. .G...AAA..
J20 I           ........T. .......A.. T..C.CT..T ..AC.G..T. .T...AGA..
MS2 I           ........C. .......G.. C..T.CT..C ..CA.A..C. .G...AAA..
M12 I           ........T. .......A.. A..A.CC..C ..CA.G..C. .T...AAA..
DL52            ........T. .......A.. T..C.AT..T ..AC.C..T. .T...CGC..
DL54            ........T. .......G.. T..T.GC..C ..AC.C..C. .C...CGT..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2460       2470       2480       2490       2500
Consensus       YRTYGGTATM GAYCTGAATG ATCARWCGAT CAAYCARCKT YTDGCWMARC
DL1 I           TG.C.....A ..C....... ....AA.... ...T..G.G. C.A..TC.A.
DL2 I           TG.C.....A ..C....... ....AA.... ...T..G.G. T.A..TC.A.
DL13 I          TG.C.....A ..C....... ....AA.... ...T..G.G. T.A..TC.A.
DL16 I          TG.C.....A ..C....... ....AA.... ...T..G.G. T.A..TC.A.
```

```
ST4 I          CG.T.....A ..T....... ....AT.... ...C..G.T. C.G..TC.G.
R17 I          CG.T.....A ..C....... ....GT.... ...C..G.G. C.A..TC.G.
J20 I          CG.T.....A ..C....... ....AA.... ...C..G.G. C.G..TC.A.
MS2 I          CG.T.....A ..C....... ....AT.... ...C..G.G. C.G..TC.G.
M12 I          CG.T.....A ..C....... ....AA.... ...C..G.G. C.G..TC.A.
DL52           TA.C.....C ..C....... ....AA.... ...T..G.G. C.T..AA.A.
DL54           TA.C.....C ..C....... ....AA.... ...T..A.G. C.T..AA.A.


                        ....|....| ....|....| ....|....| ....|....| ....|....|
                              2510       2520       2530       2540       2550
Consensus      WNGGCAGYRY MGATGGHTCD YTDGCRACKA TWGAYYTATC GTCNGCNTCB
DL1 I          AG.....CAT A.....T..T T.A..G..G. .A..CC.... ...A..T..G
DL2 I          AG.....CAT A.....C..T T.A..A..G. .A..TT.... ...G..T..T
DL13 I         AG.....CAT A.....C..T T.A..A..G. .A..TT.... ...G..T..T
DL16 I         AG.....CAT A.....C..T T.A..A..G. .A..TT.... ...G..T..T
ST4 I          AG.....CGT A.....T..G C.T..G..G. .A..CT.... ...T..A..C
R17 I          AA.....CGC A.....T..G C.T..G..G. .A..CT.... ...C..G..C
J20 I          AA.....CGT C.....C..A T.G..G..G. .A..CT.... ...G..C..T
MS2 I          AG.....CGT A.....T..G C.T..G..G. .A..CT.... ...T..A..C
M12 I          AA.....TAT A.....A..G T.A..G..T. .A..CT.... ...T..A..C
DL52           TT.....CAT C.....T..T T.A..G..G. .T..TT.... ...T..T..T
DL54           TC.....TAT C.....T..T T.A..G..G. .A..TT.... ...T..T..T


                        ....|....| ....|....| ....|....| ....|....| ....|....|
                              2560       2570       2580       2590       2600
Consensus      GAYTCYATHT CBGAYCGCCT NGTKTGGRRN TTYCTYCCAY CKSARHTRTA
DL1 I          ..T..C..C. .T..C..... G..G...AGC ..T..C...C .TG.GC.A..
DL2 I          ..T..T..C. .T..C..... G..G...AGT ..C..C...C .TG.GT.A..
DL13 I         ..T..T..C. .T..C..... G..G...AGT ..C..C...C .TG.GT.A..
DL16 I         ..T..T..C. .T..C..... G..G...AGT ..C..C...C .TG.GT.A..
ST4 I          ..T..C..C. .C..T..... G..G...AGT ..T..C...C .TG.GC.A..
R17 I          ..T..C..T. .C..C..... A..G...AAT ..C..T...C .TG.GC.A..
J20 I          ..C..C..C. .T..C..... G..G...AGT ..C..C...C .TG.GC.A..
MS2 I          ..T..C..C. .C..T..... G..G...AGT ..T..C...C .TG.GC.A..
M12 I          ..T..C..C. .T..C..... G..G...AGT ..C..C...C .TG.GC.A..
DL52           ..T..C..A. .G..C..... C..T...GAA ..C..C...T .GC.AA.G..
DL54           ..C..C..A. .G..C..... T..T...GAG ..T..C...C .GC.AA.G..


                        ....|....| ....|....| ....|....| ....|....| ....|....|
                              2610       2620       2630       2640       2650
Consensus      YKCRTAYCTB KMKMRWATYC GYTCVYMMYR HGGAATCRTW GAYGGNSRKR
DL1 I          TT.A..T..C GATCGT..C. .C..CCACTA C......G.A ..T..CGAGA
DL2 I          TT.A..T..C GATCGT..C. .C..CCACTA C......G.A ..T..GGAGA
DL13 I         TT.A..T..C GATCGT..C. .C..CCACTA C......G.A ..T..GGAGA
DL16 I         TT.A..T..C GATCGT..C. .C..CCACTA C......G.A ..T..GGAGA
ST4 I          TT.A..T..C GATCGT..C. .C..ACACTA C......G.A ..T..CGAGA
R17 I          TT.A..T..T GATCGT..C. .C..GCACTA C......G.A ..T..CGAGA
J20 I          TT.A..T..C GATCGT..C. .C..CCACTA T......A.A ..T..AGAGA
MS2 I          TT.A..T..C GATCGT..C. .C..ACACTA C......G.A ..T..CGAGA
M12 I          CT.A..T..C GATCGT..C. .C..CCACTA C......G.A ..T..TGAGA
DL52           CG.G..C..G TCGAAA..T. .T..GTCACG A......G.T ..C..ACGTG
DL54           TG.G..C..G TCTAAA..T. .T..GCCACG A......G.T ..C..ACGTA
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2660       2670       2680       2690       2700
Consensus       YRRTMSRVTG GSAMCTATTT TCCACDATGG GWAAYGGDTT YACNTTYGAR
DL1 I           CGA.ACGG.. .G.A...... .....A.... .A..T..G.. C..T..C..A
DL2 I           CGA.ACGA.. .G.A...... .....A.... .T..T..G.. C..T..C..A
DL13 I          CGA.ACGA.. .G.A...... .....A.... .T..T..G.. C..T..C..A
DL16 I          CGA.ACGA.. .G.A...... .....A.... .T..T..G.. C..T..C..A
ST4 I           CGA.ACGA.. .G.A...... .....A.... .A..T..G.. C..G..T..G
R17 I           CGA.ACGA.. .G.A...... .....G.... .A..T..G.. T..A..T..G
J20 I           CGA.ACGA.. .G.A...... .....A.... .A..T..G.. C..T..T..A
MS2 I           CGA.ACGA.. .G.A...... .....A.... .A..T..G.. C..A..T..G
M12 I           CGA.ACGA.. .G.A...... .....A.... .A..C..G.. C..T..T..G
DL52            TAG.CGAC.. .C.C...... .....A.... .T..T..T.. T..C..T..G
DL54            TGA.CGAC.. .C.C...... .....T.... .T..T..A.. T..C..T..A


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2710       2720       2730       2740       2750
Consensus       CTAGAGTCCA TGATMTTYTG GGCWATAGTN AARGCRACYM WRAYYCATTT
DL1 I           .......... ....A..T.. ...T.....G ..A..G..CC AG.TC.....
DL2 I           .......... ....C..T.. ...T.....G ..A..A..CC AG.TC.....
DL13 I          .......... ....C..T.. ...T.....G ..A..A..CC AG.TC.....
DL16 I          .......... ....C..T.. ...T.....G ..A..A..CC AG.TC.....
ST4 I           .......... ....A..C.. ...A.....C ..A..G..CC AA.TC.....
R17 I           .......... ....A..C.. ...A.....C ..A..A..CC AA.TC.....
J20 I           .......... ....A..T.. ...T.....A ..A..A..CC AG.TC.....
MS2 I           .......... ....A..C.. ...A.....C ..A..G..CC AA.TC.....
M12 I           .......... ....A..T.. ...T.....C ..A..G..CC AA.TC.....
DL52            .......... ....A..T.. ...A.....C ..G..A..TA TG.TC.....
DL54            .......... ....C..T.. ...A.....T ..G..A..CA TG.CT.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2760       2770       2780       2790       2800
Consensus       TGGTAACSYY GGAACMATWG GCATCTAYGG GGACGATATY ATATGTCCCA
DL1 I           .......GCC .....C..A. .......C.. .........T ..........
DL2 I           .......GCC .....C..A. .......C.. .........T ..........
DL13 I          .......GCC .....C..A. .......C.. .........T ..........
DL16 I          .......GCC .....C..A. .......C.. .........T ..........
ST4 I           .......GCC .....C..A. .......C.. .........T ..........
R17 I           .......GCC .....C..A. .......C.. .........C ..........
J20 I           .......GCC .....C..A. .......C.. .........T ..........
MS2 I           .......GCC .....C..A. .......C.. .........T ..........
M12 I           .......GCC .....C..A. .......C.. .........T ..........
DL52            .......CTT .....A..T. .......T.. .........C ..........
DL54            .......CTT .....A..T. .......T.. .........C ..........
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2810       2820       2830       2840       2850
Consensus       SWGAGATTGC ACCYCGTGTG CTRGAGGCDC THRSCTWCTA CGGTTTYAAA
DL1 I           GT........ ...C...... ..A.....A. .TGC..A... ......T...
DL2 I           GT........ ...C...... ..A.....A. .TGC..A... ......C...
DL13 I          GT........ ...C...... ..A.....A. .TGC..A... ......C...
DL16 I          GT........ ...C...... ..A.....A. .TGC..A... ......C...
ST4 I           GT........ ...C...... ..G.....A. .TGC..A... ......C...
R17 I           GT........ ...C...... ..A.....A. .TGC..A... ......T...
J20 I           GT........ ...C...... ..A.....G. .AGC..A... ......C...
MS2 I           GT........ ...C...... ..A.....A. .TGC..A... ......T...
M12 I           GT........ ...C...... ..A.....A. .TGC..A... ......T...
DL52            CA........ ...T...... ..A.....T. .CAG..T... ......T...
DL54            CA........ ...T...... ..A.....T. .CAG..T... ......T...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2860       2870       2880       2890       2900
Consensus       CCGAATCWBY SKAARACGTT CRTSWCVGGK CKCTTTCGCG AGWSCTGYRG
DL1 I           .......TTC GT..A..... .G.GT.G..G .T........ ..AG...CG.
DL2 I           .......TTC GT..A..... .G.GT.A..G .T........ ..AG...TG.
DL13 I          .......TTC GT..A..... .G.GT.A..G .T........ ..AG...TG.
DL16 I          .......TTC GT..A..... .G.GT.A..G .T........ ..AG...TG.
ST4 I           .......TCC GT..A..... .G.GT.C..G .T........ ..AG...CG.
R17 I           .......TTC GT..A..... .G.GT.C..G .T........ ..AG...CA.
J20 I           .......TTC GT..A..... .G.GT.A..G .T........ ..AG...TG.
MS2 I           .......TTC GT..A..... .G.GT.C..G .T........ ..AG...CG.
M12 I           .......TTC GT..A..... .G.GT.C..G .T........ ..AG...CG.
DL52            .......AGT CG..G..... .A.CA.G..T .G........ ..TC...TG.
DL54            .......AGT CG..G..... .A.CA.G..T .G........ ..TC...TG.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2910       2920       2930       2940       2950
Consensus       YGCRCAYTWY TWCSGYGGTG YYGATKKCAA ACCGWTYTAY ATCARGAAAC
DL1 I           C..G..C.TT .A.C.T.... TC...GT... ....T.T..C ....A.....
DL2 I           C..G..C.TT .A.C.T.... TC...GT... ....T.C..C ....A.....
DL13 I          C..G..C.TT .A.C.T.... TC...GT... ....T.C..C ....A.....
DL16 I          C..G..C.TT .A.C.T.... TC...GT... ....T.C..C ....A.....
ST4 I           C..G..C.TT .A.C.T.... TC...GT... ....T.T..C ....A.....
R17 I           C..G..C.TT .A.C.T.... TC...GT... ....T.T..C ....A.....
J20 I           C..G..C.TT .A.C.T.... TC...GT... ....T.T..C ....G.....
MS2 I           C..G..C.TT .A.C.T.... TC...GT... ....T.T..C ....A.....
M12 I           C..G..C.TT .A.C.T.... TC...GT... ....T.C..C ....A.....
DL52            T..A..T.AC .T.G.C.... CT...TG... ....A.T..T ....A.....
DL54            T..A..T.AC .T.G.C.... CT...TG... ....A.T..T ....A.....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       2960       2970       2980       2990       3000
Consensus       CHGTYRACAA YCTCTTYKCC STBWKKCTGW TMHTNAAYMG RCTDMGSGGN
DL1 I           .A..TG.... C.....CT.. C.TATG...A .AA.G..TC. G..TA.G..G
DL2 I           .A..TG.... T.....CT.. C.TATG...A .CA.G..TC. G..TA.G..G
DL13 I          .A..TG.... T.....CT.. C.TATG...A .CA.G..TC. G..TA.G..G
DL16 I          .A..TG.... T.....CT.. C.TATG...A .CA.G..TC. G..TA.G..G
ST4 I           .T..TG.... T.....CG.. C.TATG...A .AT.G..TC. G..AC.G..T
```

```
R17 I        .T..TG.... C.....TG.. C.GATG...A .AT.A..TC. G..AC.G..T
J20 I        .A..CG.... T.....CT.. C.TATG...A .CA.G..TC. G..TA.G..A
MS2 I        .T..TG.... T.....CG.. C.GATG...A .AT.A..TC. G..AC.G..T
M12 I        .A..TG.... C.....TT.. C.TATG...A .AC.G..TC. G..AC.G..T
DL52         .C..TA.... C.....CG.. G.CTGT...T .AC.C..CA. G..AC.C..G
DL54         .C..TA.... C.....CG.. G.CTGT...T .AC.T..CA. A..GC.C..C


             ....|....| ....|....| ....|....| ....|....| ....|....|
                     3010       3020       3030       3040       3050
Consensus    TGGGGNGTTG TSRRMGGWRT GTCAGATCCD CGCCTHTWYR AGRYKTGGRW
DL1 I        .....C.... .CGGA..TA. .........G .....T.ATA ..GTT...GT
DL2 I        .....C.... .CGGA..TA. .........G .....T.ATA ..GTT...GT
DL13 I       .....C.... .CGGA..TA. .........G .....T.ATA ..GTT...GT
DL16 I       .....C.... .CGGA..TA. .........G .....T.ATA ..GTT...GT
ST4 I        .....A.... .CGGA..TA. .........A .....T.ACA ..GTG...GT
R17 I        .....G.... .CGGA..TA. .........A .....T.ACA ..GTG...GT
J20 I        .....T.... .CGGA..TA. .........T .....A.ATA ..GTT...GT
MS2 I        .....A.... .CGGA..TA. .........A .....C.ATA ..GTG...GT
M12 I        .....G.... .CGGA..TA. .........G .....T.ACA ..GTT...GT
DL52         .....T.... .GAAC..AG. .........T .....C.TCG ..ACT...AA
DL54         .....T.... .GAAC..TG. .........T .....C.TCG ..ACT...AA


             ....|....| ....|....| ....|....| ....|....| ....|....|
                     3060       3070       3080       3090       3100
Consensus    AYGRCTVTCC KMMCDKGTDC CHTCGATRYT YTTCGGTSSS WCDRACCTYG
DL1 I        .C.A..G... TCC.TG..T. .A.....GT. C......GGG A.GG....T.
DL2 I        .C.A..A... TCC.TG..T. .A.....GT. C......GGG A.GG....T.
DL13 I       .C.A..A... TCC.TG..T. .A.....GT. C......GGG A.GG....T.
DL16 I       .C.A..A... TCC.TG..T. .A.....GT. C......GGG A.GG....T.
ST4 I        .C.A..C... TCC.AG..G. .T.....GT. T......GGG A.GG....C.
R17 I        .C.A..C... TCC.AG..A. .T.....GT. C......GGG A.GG....C.
J20 I        .C.A..G... TCC.TG..T. .A.....GT. C......GGG A.GG....T.
MS2 I        .C.G..C... TCC.AG..G. .T.....GT. C......GGG A.GG....C.
M12 I        .C.A..G... TCC.TG..T. .A.....GT. C......CCC A.GG....C.
DL52         .T.G..A... GAA.GT..T. .T.....AC. T......GGG T.TA....T.
DL54         .T.G..A... GAA.GT..T. .C.....AC. T......GGG T.AA....T.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3110       3120       3130       3140       3150
Consensus       MTGCHGACTA YTACGTRGTY AGYCSBCMNR MNGCYADRWM KCRKTDYGAB
DL1 I           C...C..... T.....A..C ..C.CC.CGA CC..--AGTC T.AG.TT-.T
DL2 I           C...C..... C.....A..T ..C.CT.CGA AT..--TGTC T.GG.TT-.T
DL13 I          C...C..... C.....A..T ..C.CT.CGA AT..--TGTC T.GG.TT-.T
DL16 I          C...C..... C.....A..T ..C.CT.CGA AT..--TGTC T.GG.TT-.T
ST4 I           C...C..... C.....A..C ..C.CG.CCA CG..--AGTC T.GG.AT-.T
R17 I           C...C..... C.....A..C ..C.CG.CTA CG..--GGTC T.GG.AT-.T
J20 I           C...T..... C.....A..C ..C.CC.CGA CA..--GGTC T.AG.TT-.T
MS2 I           C...C..... C.....A..C ..C.CG.CTA CG..--AGTC T.GG.AT-.C
M12 I           C...T..... C.....A..C ..C.CT.CAA CT..--TGTC T.AG.AT-.T
DL52            A...A..... C.....G..T ..T.GT.AGG AA..T.AAAA G.GT.AC..G
DL54            A...A..... T.....G..C ..C.GC.AGG AA..C.GGAA G.GT.GC..G


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3160       3170       3180       3190       3200
Consensus       RVYMANMCHB YNTAYGGRMG BHYVCTYKCK SAYDYYCRYA SCWCKSSTYW
DL1 I           ACTA.AA.TG CC..C..GA. GTTA..CG.G G.TACC.GT. C.T.GGG.TT
DL2 I           ACCA.AA.TG CA..T..GA. GTTA..CG.G G.CGCC.GT. C.T.GGG.TT
DL13 I          ACCA.AA.TG CA..T..GA. GTTA..CG.G G.CGCC.GT. C.T.GGG.TT
DL16 I          ACCA.AA.TG CA..T..GA. GTTA..CG.G G.CGCC.GT. C.T.GGG.TT
ST4 I           ACCA.GA.TC CG..T..GC. GCTA..CG.G G.TACC.GT. C.T.GGG.TT
R17 I           ACCA.GA.TC CG..T..AC. GCTG..CG.G G.TACC.GT. C.T.GGG.TT
J20 I           ACTA.GA.CG CA..T..GA. GTTG..CG.G G.CGCC.GT. C.T.GGG.TT
MS2 I           ACCA.GA.TC CG..C..GC. GCTG..CG.G G.TACC.GT. C.T.GGG.TT
M12 I           ACTA.GA.CG CG..T..GA. GCTG..CG.G G.TACC.GT. C.T.GGG.TT
DL52            AGTC.CC.AT CT..T..GC. CACC..CT.T C.CTTT.AT. G.A.TCC.CA
DL54            GATC.TC.AG TA..T..AC. TACC..TG.T C.TTTC.AC. G.T.TCC.CA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3210       3220       3230       3240       3250
Consensus       YCGKYTKKMY MGHRYRCCCG YNHVWRARCG HRARYDYTCT YMRCGHRAAG
DL1 I           C..TC.TGCT C.TAT---.. CAAAAG.G.. AA.GCGC.-. TAG..AG...
DL2 I           C..TC.TGCT C.TAT---.. CGAAAG.G.. AA.GCAC.-. TAG..AG...
DL13 I          C..TC.TGCT C.TAT---.. CGAAAG.G.. AA.GCAC.-. TAG..AG...
DL16 I          C..TC.TGCT C.TAC---.. CGAAAG.G.. AA.GCAC.-. TAG..AG...
ST4 I           C..TC.TGCT C.TAT---.. CTCGAG.A.. CA.GTTC.-. CAG..AA...
R17 I           C..TC.TGCT C.TAT---.. CTCGAG.A.. CA.GTTC.-. CAG..AA...
J20 I           C..TC.TGCT C.TAT---.. CGAAAG.G.. TA.GCAC.-. TAG..AG...
MS2 I           C..TC.TGCT C.TAT---.. CTCGAG.A.. CA.GTTC.-. CAG..AA...
M12 I           C..TC.TGCT A.AAT---.. CAAAAG.G.. TA.GCGC.-. TAG..AG...
DL52            T..GC.TTAC C.CGTG.... TATCTA.A.. TG.ATTT... CCA..TG...
DL54            T..GT.GTAC C.CGTA.... CCTCTA.A.. CG.ATTT... TCG..CG...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3260       3270       3280       3290       3300
Consensus       MRYGRCASYG GYCGCYWCAT MRCDTGGTWC CATAMTGGHG GTSARRTYAY
DL1 I           CAT.A..GT. .T...TA... AG.A....T. ....C...A. ..G.GA.C.C
DL2 I           CAT.A..GT. .T...TA... AG.A....T. ....C...A. ..G.AA.C.C
DL13 I          CAT.G..GT. .T...TA... AG.A....T. ....C...A. ..G.AA.C.C
DL16 I          CAT.A..GT. .T...TA... AG.A....T. ....C...A. ..G.AA.C.C
ST4 I           CAT.A..GT. .C...TA... AG.G....T. ....C...A. ..G.AG.C.C
```

```
R17 I          CAT.A..GT. .T...TA... AG.G....T. ....C...A. ..G.AA.C.C
J20 I          CAT.A..GT. .T...TA... AG.A....T. ....C...A. ..G.AA.C.C
MS2 I          CAC.A..GT. .T...TA... AG.G....T. ....C...A. ..G.AA.C.C
M12 I          CAT.A..GT. .T...TA... AG.G....T. ....C...A. ..G.AA.C.C
DL52           AAT----CC. .T...CT... CA.T....A. ....A...T. ..C.AG.T.T
DL54           AGT----CC. .T...CT... CA.T....A. ....A...C. ..C.AA.T.T


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3310       3320       3330       3340       3350
Consensus      YGAYASYAYK AMGWCCSCHR GSGTKCGYVT HVTRCGMACD TCGGARTGGC
DL1 I          C..T.GT.TG .A.T..G.TG .C..G..CG. AA.G..C..T .....G....
DL2 I          C..T.GT.TG .A.T..G.TG .C..G..TG. AA.A..C..T .....G....
DL13 I         C..T.GT.TG .A.T..G.TG .C..G..TG. AA.A..C..T .....G....
DL16 I         C..T.GT.TG .A.T..G.TG .C..G..TG. AA.A..C..T .....G....
ST4 I          C..C.GT.TG .A.T..G.CG .C..G..TA. TA.G..C..T .....G....
R17 I          C..C.GT.TG .A.T..G.CG .C..G..CA. CA.G..C..T .....G....
J20 I          C..T.GT.TG .A.T..G.CG .C..G..TG. TA.G..C..T .....A....
MS2 I          C..C.GC.TG .A.T..G.CG .C..G..CG. TA.A..C..T .....G....
M12 I          C..T.GT.TG .A.T..G.TG .C..G..CG. AC.G..C..T .....G..--
DL52           T..C.CC.CT .C.A..C.AA .G..T..CC. TG.G..A..G .....A....
DL54           T..C.CC.CT .C.A..C.AA .G..T..CC. TG.G..A..A .....A....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3360       3370       3380       3390       3400
Consensus      TRRCRSYRGT KCCMHYMTTC CCKCAGGARK RTGRCRMMHG CGAGCTCTCC
DL1 I          .GA.GCCG.. T..CACA... ..T.....GT G..G-GCCA. ..........
DL2 I          .GA.GCCG.. T..CATA... ..T.....AT G..G-GCCA. ..........
DL13 I         .GA.GCCG.. T..CATA... ..T.....AT G..G-GCCA. ..........
DL16 I         .GA.GCCG.. T..CATA... ..T.....AT G..G-GCCA. ..........
ST4 I          .AA.GCCG.. T..CACA... ..T.....GT G..G-GCCA. ..........
R17 I          .AA.GCCG.. T..CACA... ..T.....GT G..G-GCCA. ..........
J20 I          .AA.GCCG.. T..CACA... ..G.....GT G..G-GCCA. ..........
MS2 I          .AA.GCCG.. T..CACA... ..T.....GT G..G-GCCA. ..........
M12 I          ---------- ---------- ---------- ---------- ----------
DL52           .AA.AGTG.. G..ACTC... ..T.....AG A..G.AACT. ..........
DL54           .AG.GGTA.. G..ATCC... ..T.....AG A..A.ACAC. ..........

               STOP 4 Grp I      STOP 4 JS
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       3410       3420       3430       3440       3450
Consensus      TMGKYAGCWS RSCKAGGGAC CCCCGTAAWC GGGGTGGGTG TGCYCGMRAR
DL1 I          .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
DL2 I          .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
DL13 I         .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
DL16 I         .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
ST4 I          .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
R17 I          .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
J20 I          .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
MS2 I          .C.GT...TG AC.G..... ........A. .......... ...T..AA.G
M12 I          ---------- ---------- ---------- ---------- ----------
DL52           .A.TC...AC GG.T..... ........T. .......... ---C..CG.A
DL54           .A.TC...AC GG.T..... ........T. .......... ---C..CG.A
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3460       3470       3480       3490       3500
Consensus       AGCRSSKRTY CRYSDWARCR RYCCGGMTSS AYMGRARKRW BSKCSKRMTT
DL1 I           ...ACGGG.C .GCGAA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
DL2 I           ...ACGGG.C .GCGAA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
DL13 I          ...ACGGG.C .GCGAA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
DL16 I          ...ACGGG.C .GCGAA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
ST4 I           ...ACGGG.C .GCGAA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
R17 I           ...ACGGG.C .GCGAA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
J20 I           ...ACGGG.C .GCGTA.G.G GT---.C.CC .CC.A.AGGT GGG.GGGC..
MS2 I           ...ACGGG.- -GCGAA.G.G GT....C.CC .CC.A.AGGT GGG.GGGC..
M12 I           ---------- ---------- ---------- ---------- ----------
DL52            ...GGCTA.T .ATCGT.A.A AC---.A.GG .TA.G.GTAA TCT.CTAA..
DL54            ...GGCTA.T .ATCGT.A.A AC---.A.GG .TA.G.GTAA CCT.CTAA..

                  (See App C4) This 3' alignment of JS is not accurate.
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        3510       3520       3530       3540       3550
Consensus       YGGYSMMKRR MYYTSCYYBK VWAGWSASSR CCCRGGATTC TCCCGATTTG
DL1 I           C..CCCAGGG ACC.C.CCCT AA..AG.GGA ...G..... .........
DL2 I           C..CCCAGGG ACC.C.CCCT AA..AG.GGA ...G..... .........
DL13 I          C..CCCAGGG ACC.C.CCCT AA..AG.GGA ...G..... .........
DL16 I          C..CCCAGGG ACC.C.CCCT AA..AG.GGA ...G..... .........
ST4 I           C..CCCAGGG ACC.C.CCTT GA..AG.GGG ...G..... .........
R17 I           C..CCCAGGG ACC.C.CCCT GA..AG.GGA ...G..... .........
J20 I           C..CCCAGGG ACC.C.CCCT AA..AG.GGA ...G..... .........
MS2 I           C..CCCAGGG ACC.C.CCCT AA..AG.GGA ...G..... .........
M12 I           ---------- ---------- ---------- ---------- ----------
DL52            T..TGACTAA CTT.G.TTGG CT..TC.CCA ...A------ ----------
DL54            C..TGACTAA CTT.G.TTGG CT..TC.CCA ...A------ ----------


                ....|....| ....|....| ....|....|
                        3560       3570       3580
Consensus       GTAACTAGCT GCTTGGCTAG TKACCACCCA
DL1 I           .......... .......... .T........
DL2 I           .......... .......... .T........
DL13 I          .......... .......... .T........
DL16 I          .......... .......... .T........
ST4 I           .......... .......... .T........
R17 I           .......... .......... .G........
J20 I           .......... .......... .T........
MS2 I           .......... .......... .T........
M12 I           ---------- ---------- ----------
DL52            ---------- ---------- ----------
DL54            ---------- ---------- ----------
```

Appendix C2
JS Amino Acids

Amino acid sequences of *Leviviridae* Group JS. Note the YGDD motif in all *Leviviridae* replicase proteins.

Group JS isolates DL52 and DL54. The amino acid composition is highly conserved among these two strains.

Alignment: Align JS strains maturation protein.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL52 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL54 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR


                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL52 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL54 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL52 matur      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL54 matur      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL52 matur      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL54 matur      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL52 matur      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL54 matur      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL52 matur      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL54 matur      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
Consensus       LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL52 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL54 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK


                ....|....| ....|....| ....|....| ....|....| ...
                       360        370        380        390
Consensus       AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL52 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL54 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
```

Alignment: Align JS strains capsid.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
Consensus       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL52 capsi      MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL54 coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
Consensus       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL52 capsi      QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL54 coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND


                ....|....| ....|....| ....|....|
                       110        120        130
Consensus       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL52 capsi      DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL54 coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
```

Alignment: Align JS strains lysis protein.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
Consensus       METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL52 lysis      METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL54 lysis      METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK


                ....|....| ....|....| ....|
                        60         70
Consensus       FTNQLLLSLL EAVIRTVETL QQLLT
DL52 lysis      FTNQLLLSLL EAVIRTVETL QQLLT
DL54 lysis      FTNQLLLSLL EAVIRTVETL QQLLT
```

Alignment: Align JS strains replicase protein.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL52 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL54 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL52 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL54 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL52 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL54 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       QATVTPRALR AALLVRDQC  PWIRHAV Y  ESYEFRLVVG NGVFTVPKNN
DL52 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL54 repli      QATVTPRALR AALLVRDQCA PWIRHAVHYS ESYEFRLVVG NGVFTVPKNN

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       KIDRAACKEP DANMYLQKGV G FIRRRLRS IGIDLNDQTI NQRLAKLGSI
DL52 repli      KIDRAACKEP DANMYLQKGV GDFIRRRLRS IGIDLNDQTI NQRLAKLGSI
DL54 repli      KIDRAACKEP DANMYLQKGV GGFIRRRLRS IGIDLNDQTI NQRLAKLGSI

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       DGSLATIDLS SASDSISDRL VWEFLP QMY AYLSKIRS R GIVDGR  DW
DL52 repli      DGSLATIDLS SASDSISDRL VWEFLPSQMY AYLSKIRSSR GIVDGRVVDW
DL54 repli      DGSLATIDLS SASDSISDRL VWEFLPPQMY AYLSKIRSPR GIVDGRMIDW

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       HLFSTMGNGF TFELESMIFW AIVKATM HF GNLGTIGIYG DDIICPTEIA
DL52 repli      HLFSTMGNGF TFELESMIFW AIVKATMIHF GNLGTIGIYG DDIICPTEIA
DL54 repli      HLFSTMGNGF TFELESMIFW AIVKATMTHF GNLGTIGIYG DDIICPTEIA

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus       PRVLEALSFY GFKPNQSKTF ITGRFRESCG AHYFGGADCK PIYIKKPVNN
DL52 repli      PRVLEALSFY GFKPNQSKTF ITGRFRESCG AHYFGGADCK PIYIKKPVNN
DL54 repli      PRVLEALSFY GFKPNQSKTF ITGRFRESCG AHYFGGADCK PIYIKKPVNN
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus       LFAVCLLLNR LRGWGVVNGV SDPRLFETWK WLSERVPSIL FGGSNLDADY
DL52 repli      LFAVCLLLNR LRGWGVVNGV SDPRLFETWK WLSERVPSIL FGGSNLDADY
DL54 repli      LFAVCLLLNR LRGWGVVNGV SDPRLFETWK WLSERVPSIL FGGSNLDADY


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        460        470        480        490        500
Consensus       YVVSRQEA K R E HP YGR TL HFHS PH RLYRVP SKR EFS REESGR
DL52 repli      YVVSRQEAKK RYESHPSYGR TLSHFHSTPH RLYRVPVSKR EFSPREESGR
DL54 repli      YVVSRQEARK RCEDHPVYGR TLAHFHSSPH RLYRVPASKR EFSSREESGR


                ....|....| ....|....| ....|....| ....|....| ....|
                        510        520        530        540
Consensus       LITWYHNGGQ  IDTTTTPRV RLVRTSEWL  VVP FPQED    ELS
DL52 repli      LITWYHNGGQ VIDTTTTPRV RLVRTSEWLT VVPLFPQEDG NCELS
DL54 repli      LITWYHNGGQ IIDTTTTPRV RLVRTSEWLA VVPSFPQEDD TRELS
```

Appendix C3
Group I and JS Amino Acids

Amino acid sequences of *Leviviridae* Group I and JS strains (DL52, DL54). Group I strains DL1, DL2, DL13, DL16, J20, ST4, R17, M12, MS2, fr.

A. Align maturation protein.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MR F           V   R  YA G  EDNS   L YRSNW  PG   STG
DL52 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL54 MATUR      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL1 matura      MRAFSVLDKE SETFVPLVRT YADGEVEDNS FSLKYRSNWT PGRFNSTGAR
DL2 matura      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL13 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
DL16 matur      MRAFSVLDQE SETFVPSVRV YADGQVEDNS FSLKYRSNWT PGRFNSTGSR
J20 matura      MRAFSVLDRE SETFVPSVRV YADGEVEDNS FSLKYRSNWT PGRFNSTGTR
ST4 matura      MRAFSTLDRE NETFVPSVRV YADGETEDNS FSLKYRSNWT PGRFNSTGAK
R17 matura      MRAFSALDKE SKTFVPSIRV YANGETEDNS FSLKYRSNWT PGRFNSTGAR
M12 matura      MRAFSVLDQE NETFVPSVRV YADGETEDNS FSLKYRSNWT PGRFNSTGAR
MS2 matura      MRAFSTLDRE NETFVPSVRV YADGETEDNS FSLKYRSNWT PGRFNSTGAK
fr maturat      MRKFIPTERM SKSHVVSVRE YADGELEDNS LPLIYRSNWS PGQYTSTGPR


                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       T  WHYPS Y SRGA       DQG Y R G  SWGR  EE   G G S DARS
DL52 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL54 MATUR      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL1 matura      TKQWHYPSSY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GYGFSLDARS
DL2 matura      TKQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL13 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
DL16 matur      TEQWHYPSPY SRGALSVTSV DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
J20 matura      TNQWHYPSPY SRGALSVTSV DQGSYKRSGS SWGRPYEEKA GFGFSLDARS
ST4 matura      TKQWHYPSPY SRGALSVTSI DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
R17 matura      TKQWHYPSPY SRGALSVTSI DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
M12 matura      TKQWHYPSPY SRGALSVTAI DQGAYKRSGS SWGRPYEEKT GFGFSLDARS
MS2 matura      TKQWHYPSPY SRGALSVTSI DQGAYKRSGS SWGRPYEEKA GFGFSLDARS
fr maturat      TKEWHYPSSY SRGAIGIKAL DQGKYARLGT SWGREFEERA GYGMSIDARS


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       CYSLFPVSQN  T I VP NV ANRA TEVL  KVTQGNFNLG VALAEARSTA
DL52 matur      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL54 MATUR      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL1 matura      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL2 matura      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL13 matur      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
DL16 matur      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
J20 matura      CYSLFPVSQN MTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
ST4 matura      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
R17 matura      CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
M12 matura      CYSLFPVSQN LTYIEVPQNV ANRATTEVLQ KVTQGNFNLG VALAEARSTA
```

MS2 matura CYSLFPVSQN LTYIEVPQNV ANRASTEVLQ KVTQGNFNLG VALAEARSTA
fr maturat CYSLFPVSQN LTWIDVPTNV ANRATTEVLG KVTQGNFNLG VALAEARSTA

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
Consensus       SQL TQTIAL  KAYTAARRG NWRQ  RYLA LNE RKF SK  VA RWLELQ
DL52 matur      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL54 MATUR      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL1 matura      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL2 matura      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL13 matur      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
DL16 matur      SQLATQTIAL VKAYTAARRG NWRQTLRYLA LNEDRKFRSK HVAGRWLELQ
J20 matura      SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
ST4 matura      SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
R17 matura      SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
M12 matura      SQLATQTIAL VKAYTAARRG NWRQPVRYLA LNEDRKFRSK HVAGRWLELQ
MS2 matura      SQLATQTIAL VKAYTAARRG NWRQALRYLA LNEDRKFRSK HVAGRWLELQ
fr maturat      SQLSTQTIAL IKAYTAARRG NWRQALRYLA LNENRKFNSK SVASRWLELQ

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
Consensus       FGW PL SDI QGAYEMLTKV HL  F PMRA VRQVG N  L  GRL  PAA
DL52 matur      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL54 MATUR      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL1 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNVKL DGRLSYPAAN
DL2 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL13 matur      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
DL16 matur      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
J20 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNVKL DGRLSYPAAN
ST4 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL NGRLSYPAAN
R17 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLAYPAAN
M12 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
MS2 matura      FGWLPLMSDI QGAYEMLTKV HLQEFLPMRA VRQVGTNIKL DGRLSYPAAN
fr maturat      FGWMPLLSDI QGAYEMLTKV HLKAFMPMRA VRQVGQNVSL SGRLTSPAAS

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
Consensus          TCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSF VDWL
DL52 matur      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL54 MATUR      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL1 matura      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL2 matura      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL13 matur      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
DL16 matur      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
J20 matura      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
ST4 matura      FQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
R17 matura      FQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
M12 matura      YQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
MS2 matura      FQTTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFVVDWL
fr maturat      YKSTCNISRR IVIWFYINDA RLAWLSSLGI LNPLGIVWEK VPFSFLVDWL
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       LPVGNMLEGL TAP GCSY S GTVTDVI GE S I  D  YG W   R  TAK
DL52 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL54 MATUR      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL1 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
DL2 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL13 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
DL16 matur      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVDRQGTAK
J20 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
ST4 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
R17 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
M12 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
MS2 matura      LPVGNMLEGL TAPVGCSYMS GTVTDVITGE SIISVDAPYG WTVERQGTAK
fr maturat      LPVGNMLEGL TAPIGCSYQS GTVTDVISGE STITADDIYG WDTVRPATAK


                ....|....| ....|....| ....|....| ....|....| ...
                        360        370        380        390
Consensus          SA HRGV QSV PTTG Y VKSPFS VHT LDALAL RQR L
DL52 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL54 MATUR      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL1 matura      AQISAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LSR
DL2 matura      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL13 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
DL16 matur      AQVSAMHRGV QSVWPTTGVY VKSPFSMVHT LDALALIRQR LLR
J20 matura      AQISAMHRGV QSVWPTTGVY VKSPFSIVHT LDALALIRQR LSR
ST4 matura      AQISAMHRGV QSVWPTTGAY VKSPFSMVHT LDALALIRQR LSR
R17 matura      AHVSAMHRGV QSVWPTTGAY VKSPFSMVHT LDALALIRQR LSK
M12 matura      AQVSAMHRGV QSVCPTTGVY VKSPFSMVHT LDALALIRQR LSK
MS2 matura      AQISAMHRGV QSVWPTTGAY VKSPFSMVHT LDALALIRQR LSR
fr maturat      VQISAVHRGV QSVWPTTGVY VKSPFSMVHT LDALALFRQR LWK
```

B. Alignment of capsid proteins.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50
Consensus       MASNF  FVL VDNGGTGDV  V PSNFANGV AEWISSNSRS QAYKVTCSVR
DL52 capsi      MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL54 coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL1 coat p      MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL2 coat        MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL13 coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
DL16 coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
J20  coat       MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
ST4 coat p      MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
R17 coat p      MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
M12 coat        MASNFTQFVL VDNGGTGDVT VXPSNFANGV AEWISSNSRS QAYKVTCSVR
MS2 coat p      MASNFTQFVL VDNGGTGDVT VAPSNFANGV AEWISSNSRS QAYKVTCSVR
fr coat pr      MASNFEEFVL VDNGGTGDVK VAPSNFANGV AEWISSNSRS QAYKVTCSVR


                ....|....| ....|....| ....|....| ....|....| ....|....|
                          60         70         80         90        100
Consensus       QSSA NRKYT  KVEVPKVAT Q  GGV LPV AAWRSY NME LTIP FATN
DL52 capsi      QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL54 coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL1 coat p      QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL2 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL13 coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
DL16 coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
J20  coat       QSSAQNRKYT IKVEVPKVAT QTVGGVQLPV AAWRSYLNME LTIPIFATND
ST4 coat p      QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
R17 coat p      QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
M12 coat        QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
MS2 coat p      QSSAQNRKYT IKVEVPKVAT QTVGGVELPV AAWRSYLNME LTIPIFATNS
fr coat pr      QSSANNRKYT VKVEVPKVAT QVQGGVELPV AAWRSYMNME LTIPVFATND


                ....|....| ....|....| ....|....|
                         110        120        130
Consensus       DC LIVKA Q G  K GNPI   AIAANSGIY
DL52 capsi      DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL54 coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL1 coat p      DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL2 coat        DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL13 coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
DL16 coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
J20  coat       DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
ST4 coat p      DCELIVKAMQ GLLKDGNPIP SAIAANSGIY
R17 coat p      DCELIVKAMQ GLLKDGNPIP SAIAANSGIY
M12 coat        DCALIVKAMQ GLLKDGNPIP SAIAANSGIY
MS2 coat p      DCELIVKAMQ GLLKDGNPIP SAIAANSGIY
fr coat pr      DCALIVKALQ GTFKTGNPIA TAIAANSGIY
```

C. Alignment of lysis proteins.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       M       SQ  T  S      PF HE YPC   QQRSSTLYV LI LAIFLSK
DL52 lysis      METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL54 lysis      METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL2 lysis       METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL13 lysis      METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL16 lysis      METRSPQQSQ QTPESTNRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
J20 lysis       METQSPQQSQ PTPESINRFR PFQHEDYPCR RQQRSSTLYV LIFLAIFLSK
DL1 lysis       METRSPQQSQ QTPESTNRFR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
ST4 lysis       METRFPQQSQ QTPASTNRRR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
R17 lysis       METRFPQQSQ QTPASTNRCR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
M12 lysis       METRFLRQSQ QTPASTNRYR PFKHEDYPCR XQQRSSTLYV LIFLAIFLSK
MS2 lysis       METRFPQQSQ QTPASTNRRR PFKHEDYPCR RQQRSSTLYV LIFLAIFLSK
fr lysis p      MQ----QPSQ PTRESTKKPV PFQHEEYPCQ NQQRSSTLYV LICLAIFLSK


                ....|....| ....|....| ....|
                         60         70
Consensus       FTNQLL SLL    IR V T   QLLT
DL52 lysis      FTNQLLLSLL EAVIRTVETL QQLLT
DL54 lysis      FTNQLLLSLL EAVIRTVETL QQLLT
DL2 lysis       FTNQLLLSLL EAVIRTVETL QQLLT
DL13 lysis      FTNQLLLSLL EAVIRTVETL QQLLT
DL16 lysis      FTNQLLLSLL EAVIRTVETL QQLLT
J20 lysis       FTNQLLLSLL EAVIRTVETL RQLLT
DL1 lysis       FTNQLLLSLL EAVIRTVETL QQLLT
ST4 lysis       FTNQLLLSLL EAVIRTVTTL QQLLT
R17 lysis       FTNQLLLSLL EAVIRTVTTL QQLLT
M12 lysis       FTNQLLLSLL DAVIRTVTTF QQLLT
MS2 lysis       FTNQLLLSLL EAVIRTVTTL QQLLT
fr lysis p      FTNQLLASLL DLLIRIVTTL QQLLT
```

D. Alignment of replicase.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
Consensus       MSK TKKFNS LCIDL  DLS LE YQSIASV ATGS  PHS  DFTAIAYLRD
DL52 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL54 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL1 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL2 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL13 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL16 repli      MSKTTKKFNS LCIDLPCDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
ST4 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSD DFTAIAYLRD
R17 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGNPHSD DFTAIAYLRD
J20 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
MS2 replic      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSD DFTAIAYLRD
fr replica      MSKSTKKFNS LCIDLSRDLS LEVYQSIASV ATGSSDPHSD DFTAIAYLRD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
Consensus       ELLTKHP LG  GNDEATRR  LAIAKL EAN    GQINR G FLHD   SWD
DL52 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL54 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL1 replic      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL2 replic      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL13 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL16 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
ST4 replic      ELLTKHPTLG SGNDEATRRT LAIAKLREAN DRCGQINREG FLHDKSLSWD
R17 replic      ELLTKHPTLG SGNDEATRRT LAIAKLREAN DRCGQINREG FLHDKSLSWD
J20 replic      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
MS2 replic      ELLTKHPTLG SGNDEATRRT LAIAKLREAN GDRGQINREG FLHDKSLSWD
fr replica      ELLTKHPNLG DGNDEATRRS LAIAKLLEAN DRCGQINRDG FLHDATASWD

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
Consensus       PDVLQTSIRS LIGNLLSGY  S LF  CTFS NGA MGHKLQ DAAPYKKFAE
DL52 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL54 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL1 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL2 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL13 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL16 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
ST4 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
R17 replic      PDVLQTSIRS LIGNLLSGYQ SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
J20 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
MS2 replic      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGAPMGHKLQ DAAPYKKFAE
fr replica      PDVLQTSIRS LIGNLLSGYS SQLFRHCTFS NGASMGHKLQ DAAPYKKFAE
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       QATVTPRAL  AA LV DQC  PWIRH      ESY FRLV G NGVFTVPKNN
DL52 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL54 repli      QATVTPRALR AALLVRDQCA PWIRHAVHYS ESYEFRLVVG NGVFTVPKNN
DL1 replic      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL2 replic      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL13 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
DL16 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
ST4 replic      QATVTPRALR AALLVRDQCA PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
R17 replic      QATVTPRALR AALLVRDQCA PWIRHAVHYN ESYEFRLVVG NGVFTVPKNN
J20 replic      QATVTPRALR AALLVRDQCA PWIRHAVRYN ESYKFRLVVG NGVFTVPKNN
MS2 replic      QATVTPRALR AALLVRDQCA PWIRHAVRYN ESYEFRLVVG NGVFTVPKNN
fr replica      QATVTPRALK AAVLVKDQCS PWIRHSHVFP ESYTFRLVGG NGVFTVPKNN

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       KIDRAACKEP D NMYLQKGV G FIRRRL    GIDLNDQ I NQ LA  GS
DL52 repli      KIDRAACKEP DANMYLQKGV GDFIRRRLRS IGIDLNDQTI NQRLAKLGSI
DL54 repli      KIDRAACKEP DANMYLQKGV GGFIRRRLRS IGIDLNDQTI NQRLAKLGSI
DL1 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
DL2 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
DL13 repli      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
DL16 repli      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQTI NQRLAQQGSI
ST4 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQSI NQLLAQQGSV
R17 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQSI NQRLAQQGSA
J20 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLRS VGIDLNDQTI NQRLAQQGSV
MS2 replic      KIDRAACKEP DMNMYLQKGV GAFIRRRLKS VGIDLNDQSI NQRLAQQGSV
fr replica      KIDRAACKEP DMNMYLQKGV GGFIRRRLKT VGIDLNDQTI NQRLAQQGSR

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       DGSLATIDLS SASDSISDRL VW FLP   Y  YL  IRS   G   G    W
DL52 repli      DGSLATIDLS SASDSISDRL VWEFLPSQMY AYLSKIRSSR GIVDGRVVDW
DL54 repli      DGSLATIDLS SASDSISDRL VWEFLPPQMY AYLSKIRSPR GIVDGRMIDW
DL1 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
DL2 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
DL13 repli      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
DL16 repli      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
ST4 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
R17 replic      DGSLATIDLS SASDSISDRL VWNFLPPELY SYLDRIRSHY GIVDGETIRW
J20 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIIDGETIRW
MS2 replic      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDRIRSHY GIVDGETIRW
fr replica      DGSLATIDLS SASDSISDRL VWSFLPPELY SYLDMIRSHY GYVNGKMIRW
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus        LFSTMGNGF TFELESMIFW AIV AT  HF  N GTIGIYG DDIICP EIA
DL52 repli      HLFSTMGNGF TFELESMIFW AIVKATMIHF GNLGTIGIYG DDIICPTEIA
DL54 repli      HLFSTMGNGF TFELESMIFW AIVKATMTHF GNLGTIGIYG DDIICPTEIA
DL1 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
DL2 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
DL13 repli      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
DL16 repli      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
ST4 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
R17 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
J20 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
MS2 replic      ELFSTMGNGF TFELESMIFW AIVKATQIHF GNAGTIGIYG DDIICPSEIA
fr replica      ELFSTMGNGF TFELESMIFW AIVRATQIHF RNTGTIGIYG DDIICPTEIA


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus       PRVLEAL  Y GFKPN  KTF   G FRESC  AH   G D K P YI KP
DL52 repli      PRVLEALSFY GFKPNQSKTF ITGRFRESCG AHYFGGADCK PIYIKKPVNN
DL54 repli      PRVLEALSFY GFKPNQSKTF ITGRFRESCG AHYFGGADCK PIYIKKPVNN
DL1 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
DL2 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
DL13 repli      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
DL16 repli      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
ST4 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
R17 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCS AHFYRGVDVK PFYIKKPVDN
J20 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIRKPVDN
MS2 replic      PRVLEALAYY GFKPNLRKTF VSGLFRESCG AHFYRGVDVK PFYIKKPVDN
fr replica      PRVLEALSFY GFKPNLRKTF TSGSFRESCG AHYFRGVDVK PFYIKKPITD


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus       LF   L  NR  RGWGVV G   DPRL   W   LS  VP    FGG  L ADY
DL52 repli      LFAVCLLLNR LRGWGVVNGV SDPRLFETWK WLSERVPSIL FGGSNLDADY
DL54 repli      LFAVCLLLNR LRGWGVVNGV SDPRLFETWK WLSERVPSIL FGGSNLDADY
DL1 replic      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
DL2 replic      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
DL13 repli      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
DL16 repli      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
ST4 replic      LFALMLILNR LRGWGVVGGM SDPRLYKVWV RLSSQVPSMF FGGTDLAADY
R17 replic      LFALMLILNR LRGWGVVGGM SDPRLYKVWV RLSSQVPSMF FGGTDLAADY
J20 replic      LFSLMLIMNR LRGWGVVGGM SDPRLYKVWV RLSSLVPSMF FGGTDLAADY
MS2 replic      LFALMLILNR LRGWGVVGGM SDPRLYKVWV RLSSQVPSMF FGGTDLAADY
fr replica      LFSLMLILNR IRGWGVVNGI ADPRLYEVWE KLSRLVPRYL FGGTDLQADY


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        460        470        480        490        500
Consensus       YVVS               R             L R        FS    SGR
DL52 repli      YVVSRQEAKK RYESHPSYGR TLSHFHSTPH RLYRVPVSKR EFSPREESGR
DL54 repli      YVVSRQEARK RCEDHPVYGR TLAHFHSSPH RLYRVPASKR EFSSREESGR
DL1 replic      YVVSPPTAVS VYTKTA-YGR LLADTRTSGF RLARIAKERK RFSEKHDSGR
DL2 replic      YVVSPPNAVS VYTKTA-YGR LLADARTSGF RLARIAKERK HFSEKHDSGR
DL13 repli      YVVSPPNAVS VYTKTA-YGR LLADARTSGF RLARIAKERK HFSEKHGSGR
```

```
DL16 repli   YVVSPPNAVS VYTKTA-YGR LLADARTSGF RLARTAKERK HFSEKHDSGR
ST4 replic   YVVSPPTAVS VYTKTP-YGR LLADTRTSGF RLARIARERK FFSEKHDSGR
R17 replic   YVVSPPTAVS VYTKTP-YGR LLADTRTSGF RLARIARERK FFSEKHDSGR
J20 replic   YVVSPPTAVS VYTKTA-YGR LLADARTSGF RLARIAKERK HFSEKHDSGR
MS2 replic   YVVSPPTAVS VYTKTP-YGR LLADTRTSGF RLARIARERK FFSEKHDSGR
fr replica   YVVSPPILKG IYSKMN-GRR EYAEARTTGF KLARIARWRK HFSDKHDSGR


             ....|....| ....|....| ....|....| ....|....| ....|.
                     510        520        530        540
Consensus     I W H GG    D      V R  RTSEWL   VP FPQE
DL52 repli   LITWYHNGGQ VIDTTTTPRV RLVRTSEWLT VVPLFPQEDG NCELS-
DL54 repli   LITWYHNGGQ IIDTTTTPRV RLVRTSEWLA VVPSFPQEDD TRELS-
DL1 replic   YIAWFHTGGE ITDSMKSAGV RVMRTSEWLT PVPTFPQECG PASSPR
DL2 replic   YIAWFHTGGE ITDSMKSAGV RVIRTSEWLT PVPIFPQECG PASSPR
DL13 repli   YIAWFHTGGE ITDSMKSAGV RVIRTSEWLT PVPIFPQECG PASSPR
DL16 repli   YIAWFHTGGE ITDSMKSAGV RVIRTSEWLT PVPIFPQECG PASSPR
ST4 replic   YIAWFHTGGE VTDSMKSAGV RIMRTSEWLT PVPTFPQECG PASSPR
R17 replic   YIAWFHTGGE ITDSMKSAGV RIMRTSEWLT PVPTFPQECG PASSPR
J20 replic   YIAWFHTGGE ITDSMKSAGV RVMRTSEWLT PVPTFPQECG PASSPR
MS2 replic   YIAWFHTGGE ITDSMKSAGV RVIRTSEWLT PVPTFPQECG PASSPR
fr replica   YIAWFHTGGE ITDSMKSAGV RVMRTSEWLQ PVPVFPQECG PASSPQ
```

Alignment: Group II replicase and JS strains.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
Consensus       M          LC D  RD           S    GS DP S  DF   AYLRD
DL52 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL54 repli      MSKTTKKFNS LCIDLPRDLS LEIYQSIASV ATGSGDPHSR DFTAIAYLRD
DL10 repli      MFRFTEIEKT LCMDRTRDCA VRFHVYLQSL DMGSSDPHSP DFDGLAYLRD
DL20 repli      MFRFTEIEKT LCMDRTRDCA VRFHVYLQSL DLGSSDPHSP DFDGLAYLRD
GA replica      MFRFREIEKT LCMDRTRDCA VRFHVYLQSL DLGSSDPLSP DFDGLAYLRD
T72 replic      MFRFTEIKKT LCMDRTRDCA VRFHVYLQSL DLGSSDPHSP DFDGLAYLRD
KU1 replic      MFRFTEIEKT LCMDRTRDCA VRFHVYLQSL DLGSSDPHSP DFDGLAYLRD


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
Consensus       E LTKHPSLG   N  A R   LA AKL      RC   N  G
DL52 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL54 repli      ELLTKHPSLG NGNDEATRRA LAIAKLREAN ERCGQINREG FLHDKSLSWD
DL10 repli      ECLTKHPSLG NSNSDARRKE LAYAKLMDSD QRCKIQNSNG YDYSHIESSV
DL20 repli      ECLTKHPSLG DSNSDARRKE LAYAKLMDSD QRCKIQNSNG YDYSHIESGV
GA replica      ECLTKHPSLG DSNSDARRKE LAYAKLMDSD QRCKIQNSNG YDYSHIESGV
T72 replic      ECLTKHPSLG DSNSDALRKE LAYAKLMDSD QRCKIQNSNG YDLSHIDSGV
KU1 replic      ECLTKHPSLG DSNSDALRKE LAYAKLMDSD QRCKIQNSNG YDLSHIDAGV


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
Consensus                    LL G   S     C FS NGAS G KL  DAAP KK A
DL52 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL54 repli      PDVLQTSIRS LIGNLLSGYR SSLFGQCTFS NGASMGHKLQ DAAPYKKFAE
DL10 repli      LSGILKTAQA LVANLLTGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG
DL20 repli      LSGILKTAQA LVANLLTGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG
GA replica      LSGILKTAQA LVANLLTGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG
T72 replic      LNGILLTAKA SIAKLLMGFE SHFLNDCSFS NGASQGFKLQ DAAPFKKIAG
KU1 replic      LNGILLTAKA LIAKLLIGFE SHFLNDCSFS NGASQGFKLR DAAPFKKIAG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
Consensus       QATVT  A    A      C  PW           E   FR V  GNG F VPK
DL52 repli      QATVTPRALR AALLVRDQCV PWIRHAVRYN --ESYEFRLV VGNGVFTVPK
DL54 repli      QATVTPRALR AALLVRDQCA PWIRHAVHYS --ESYEFRLV VGNGVFTVPK
DL10 repli      QATVTAPAYN IAVAAVKTCA PWYAYMQETY GDETRWFRRV YGNGLFSVPK
DL20 repli      QATVTAPAYD IAVAAVKTCA PWYAYMQETY GDETKWFRRV YGNGLFSVPK
GA replica      QATVTAPAYD IAVAAVKTCA PWYAYMQETY GDETKWFRRV YGNGLFSVPK
T72 replic      QATVTAPAYD LAVHAVKTCG PWLRYMQETY GDETRWFRRV YGNGLFSVPK
KU1 replic      QATVTAPAYD LAVLAVKTCG PWLRYMQETY GDETRWFRRV YGNGLFSVPK
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       NNKIDRAACK EPD NMYLQK G G FIR RL RS  IDLNDQ T NQ LA LG
DL52 repli      NNKIDRAACK EPDANMYLQK GVGDFIRRRL RSIGIDLNDQ TINQRLAKLG
DL54 repli      NNKIDRAACK EPDANMYLQK GVGGFIRRRL RSIGIDLNDQ TINQRLAKLG
DL10 repli      NNKIDRAACK EPDMNMYLQK GAGSFIRKRL RSVGIDLNDQ TCNQELARLG
DL20 repli      NNKIDRAACK EPDMNMYLQK GAGSFIRKRL RSVGIDLNDQ TRNQELARLG
GA replica      NNKIDRAACK EPDMNMYLQK GAGSFIRKRL RSVGIDLNDQ TRNQELARLG
T72 replic      NNKIDRAACK EPDMNMYLQK GAGSFIRRRL RSVNIDLNDQ TRNQELARLG
KU1 replic      NNKIDRAACK EPDMNMYLQK GAGSFIRRRL RSVNIDLNDQ TRNQELARLG


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       SIDGSLATID LSSASDS SD RLVW  LP    Y YL  IR       DG
DL52 repli      SIDGSLATID LSSASDSISD RLVWEFLPSQ MYAYLSKIRS SRGIVDGRVV
DL54 repli      SIDGSLATID LSSASDSISD RLVWEFLPPQ MYAYLSKIRS PRGIVDGRMI
DL10 repli      SIDGSLATID LSSASDSVSD RLVWDLLPPH VYSYLARIRS SFTMIDGRLH
DL20 repli      SIDGSLATID LSSASDSISD RLVWDLLPPH VYSYLARIRS SFTMIDGRLH
GA replica      SIDGSLATID LSSASDSISD RLVWDLLPPH VYSYLARIRT SFTMIDGRLH
T72 replic      SIDGSLATID LSSASDSVSD RLVWDLLPPH VYSYLHRIRS SFTMIDGQLH
KU1 replic      SIDGSLATID LSSASDSVSD RLVWDLLPPH VYSYLHRIRS SFTMIDGRLH


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus        W LFSTMGN GFTFELESMI FWA     M    G  G  G  YGDDII P
DL52 repli      DWHLFSTMGN GFTFELESMI FWAIVKATMI HFGNLGTIGI YGDDIICPTE
DL54 repli      DWHLFSTMGN GFTFELESMI FWAIVKATMT HFGNLGTIGI YGDDIICPTE
DL10 repli      KWGLFSTMGN GFTFELESMI FWALSKSVML SMGVTGSLGI YGDDIIVPVE
DL20 repli      KWGLFSTMGN GFTFELESMI FWALSKSVML SMGVTGSLGV YGDDIIVPVE
GA replica      KWGLFSTMGN GFTFELESMI FWALSKSIML SMGVTGSLGI YGDDIIVPVE
T72 replic      KWNLFSTMGN GFTFELESMI FWALSKSVMS YLGVTGLLGI YGDDIIVPTK
KU1 replic      KWNLFSTMGN GFTFELESMI FWALSNTVMS YLGVTGLLGI YGDDIIVPTK


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        360        370        380        390        400
Consensus         P  L  LS    F PN  K TF TG FRES CGAH F  A   KP Y K P
DL52 repli      IAPRVLEALS FYGFKPNQSK TFITGRFRES CGAHYFGGAD CKPIYIKKPV
DL54 repli      IAPRVLEALS FYGFKPNQSK TFITGRFRES CGAHYFGGAD CKPIYIKKPV
DL10 repli      CAPTLLKVLS AVNFLPNKKK TFTTGYFRES CGAHFFKGAD MKPFYCKRPM
DL20 repli      CAPTLLKVLS AVNFLPNQKK TFTTGYFRES CGAHFFKGAD MKPFYCKRPM
GA replica      CRPTLLKVLS AVNFLPNEEK TFTTGYFRES CGAHFFKDAD MKPFYCKRPM
T72 replic      CAPLLLQVLS AVNFLPNQKK TFTTGYFRES CGAHFFKGAS VKPFYCKRPM
KU1 replic      CAPLLLQVLS AVNFLPNQKK TFTTGYFRES CGAHFFKGAS VKPFYCKRPM


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        410        420        430        440        450
Consensus         L    LL  NR RGW     G SDPRLF   WK      P     GG NLD
DL52 repli      NNLFAVCLLL NRLRGWGVVN GVSDPRLFET WKWLSERVPS ILFGGSNLDA
DL54 repli      NNLFAVCLLL NRLRGWGVVN GVSDPRLFET WKWLSERVPS ILFGGSNLDA
DL10 repli      ETLPDVMLLC NRIRGWQTVG GMSDPRLFPI WKEFADMIPP KFKGGCNLDR
DL20 repli      ETLPDVMLLC NRIRGWQTVG GMSDPRLFPI WKEFADMIPP KFKGGCNLDR
GA replica      ETLPDVMLLC NRIRGWQTVG GMSDPRLFPI WKEFADMIPP KFKGGCNLDR
```

```
T72 replic   ETLPDIMLLC NRIRGWGTIG GISDPRLFPI WKEFADMIPP KFKGGCNLDR
KU1 replic   ETLPDVLLLC NRIRGWGTIG GISDPRLFPI WKEFADMIPP KFKGGCNLDR

             ....|....| ....|....| ....|....| ....|....| ....|....|
                     460        470        480        490        500
Consensus    D Y VS               G  L                    F    E
DL52 repli   DYYVVSRQEA KKRYESHPSY GRTLSHFHST PHRLYRVPVS KREFSPREES
DL54 repli   DYYVVSRQEA RKRCEDHPVY GRTLAHFHSS PHRLYRVPAS KREFSSREES
DL10 repli   DTYLVSPDKP ---------- GVTLVRVAKV RSGFN----- -HSFPYGHEN
DL20 repli   DTYLVSPDKP ---------- GVTLVRIAKV RSGFN----- -HAFPYGYEN
GA replica   DTYLVSPDKP ---------- GVSLVRIAKV RSGFN----- -HAFPYGHEN
T72 replic   DTYLVSPDKP ---------- GKTLVRVAKK RSGFN----- -HKFRSDYEN
KU1 replic   DTYLVSPDKP ---------- GVTLVRDATV RSGFN----- -YKFRRRQEN

             ....|....| ....|....| ....|....| ....|....| ....|...
                     510        520        530        540
Consensus    GR   W H G  G    T     R R    SE W    P FPQ E     LS
DL52 repli   GRLITWYHNG -GQVIDTTTT PRVRLVRTSE WLTVVPLFPQ EDGNCELS
DL54 repli   GRLITWYHNG -GQIIDTTTT PRVRLVRTSE WLAVVPSFPQ EDDTRELS
DL10 repli   GRYVHWLHMG SGEVLETISS ARFRCKPNSE WRTQIPLFPQ ELEACVLS
DL20 repli   GRYVHWLHMG SGEVLETISS ARFRCKPNSE WRTQIPLFPQ ELEACVLS
GA replica   GRYVHWLHMG SGEVLETISS ARYRCKPNSE WRTQIPLFPQ ELEACVLS
T72 replic   GRYIHWLHMG SGEVLETISS ARFRCKPNSE WRTQIPLFPQ EIEACVLS
KU1 replic   GRYIHWLHMG SGEVSETISS ARFRCKPNSE WRIQIPLFPQ EVEACVLS
```

Appendix C4
Group I and JS Replicase Nucleotides

Alignment: Group I and JS replicase nucleotides.

```
            ORF4 START
             ....|....| ....|....| ....|....| ....|....| ....|....|
                      10         20         30         40         50
Consensus    ATGTCGAAGA CAACAAAGAA GTTCAACTCT TTATGTATTG ATCTTCCTYG
DL1 I        .......... .......... .......... .......... .........C.
DL2 I        .......... .......... .......... .......... .........C.
DL13 I       .......... .......... .......... .......... .........C.
DL16 I       .......... .......... .......... .......... .........T.
ST4 I        .......... .......... .......... .......... .........C.
R17 I        .......... .......... .......... .......... .........C.
J20 I        .......... .......... .......... .......... .........C.
MS2 I        .......... .......... .......... .......... .........C.
DL52         .......... .......... .......... .......... .........C.
DL54         .......... .......... .......... .......... .........C.


             ....|....| ....|....| ....|....| ....|....| ....|....|
                      60         70         80         90        100
Consensus    CGATCTTTCT CTCGAAATTT ACCAATCAAT TGCTTCTGTC GCTACTGGAA
DL1 I        .......... .......... .......... .......... ..........
DL2 I        .......... .......... .......... .......... ..........
DL13 I       .......... .......... .......... .......... ..........
DL16 I       .......... .......... .......... .......... ..........
ST4 I        .......... .......... .......... .......... ..........
R17 I        .......... .......... .......... .......... ..........
J20 I        .......... .......... .......... .......... ..........
MS2 I        .......... .......... .......... .......... ..........
DL52         .......... .......... .......... .......... ..........
DL54         .......... .......... .......... .......... ..........


             ....|....| ....|....| ....|....| ....|....| ....|....|
                     110        120        130        140        150
Consensus    GCGGTRATCC GCACAGTRRM GACTTTACRG CAATTGCTTA CYTAAGRGAY
DL1 I        .....G.... .......AGA ........A. .......... .C....G..C
DL2 I        .....G.... .......AGA ........A. .......... .C....A..C
DL13 I       .....G.... .......AGA ........A. .......... .C....A..C
DL16 I       .....G.... .......AGA ........A. .......... .C....A..C
ST4 I        .....G.... .......GAC ........A. .......... .T....G..C
R17 I        .....A.... .......GAC ........A. .......... .C....G..C
J20 I        .....G.... .......AGA ........G. .......... .C....A..C
MS2 I        .....G.... .......GAC ........A. .......... .T....G..C
DL52         .....G.... .......AGA ........A. .......... .C....A..C
DL54         .....G.... .......AGA ........A. .......... .C....A..T
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
Consensus       GARTTGCTHA CWAAGCATCC VWCHTTRGGH WMTGGTAATG ACGAGGCGAC
DL1 I           ..G.....A. .T........ CT.C..A..C AA........ ..........
DL2 I           ..A.....A. .T........ GT.C..A..A AA........ ..........
DL13 I          ..A.....A. .T........ GT.C..A..A AA........ ..........
DL16 I          ..A.....A. .T........ GT.C..A..A AA........ ..........
ST4 I           ..A.....C. .A........ GA.C..A..T TC........ ..........
R17 I           ..A.....T. .A........ GA.T..A..C TC........ ..........
J20 I           ..G.....C. .A........ GT.T..A..A AA........ ..........
MS2 I           ..A.....C. .A........ GA.C..A..T TC........ ..........
DL52            ..A.....A. .T........ GT.C..A..A AA........ ..........
DL54            ..G.....T. .T........ AT.A..G..C AA........ ..........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
Consensus       CCGYCGNRCY YTAGCTATYG CTAAGCTNCG GGAGGCGAAT GRDSRDYGYG
DL1 I           ...C..AG.T T.......T. .......T.. .......... .AACGAT.T.
DL2 I           ...C..GG.T C.......C. .......C.. .......... .AGCGGT.C.
DL13 I          ...C..GG.T C.......C. .......C.. .......... .AGCGGT.C.
DL16 I          ...C..GG.T C.......C. .......C.. .......... .AGCGGT.C.
ST4 I           ...T..TA.C T.......C. .......A.. .......... .ATCGGT.C.
R17 I           ...T..CA.C T.......T. .......G.. .......... .ATCGGT.C.
J20 I           ...C..GG.T T.......C. .......C.. .......... .AACGAT.T.
MS2 I           ...T..TA.C T.......C. .......A.. .......... .GTGATC.C.
DL52            ...C..GG.T C.......C. .......C.. .......... .AGCGGT.C.
DL54            ...C..GG.T C.......T. .......C.. .......... .AGCGAT.T.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
Consensus       GYCAGATHAA YAGRGARGGT TTCTTACAYG AYAAATCCTT RTCRTGGGAT
DL1 I           .C.....C.. T..A..G... ........T. .C........ G..A......
DL2 I           .C.....T.. T..G..A... ........C. .T........ A..G......
DL13 I          .C.....T.. T..G..A... ........C. .T........ A..G......
DL16 I          .C.....T.. T..G..A... ........C. .T........ A..G......
ST4 I           .T.....A.. T..A..A... ........T. .C........ G..A......
R17 I           .C.....A.. T..A..A... ........T. .C........ G..A......
J20 I           .C.....T.. C..A..A... ........T. .C........ G..G......
MS2 I           .T.....A.. T..A..A... ........T. .C........ G..A......
DL52            .C.....T.. T..G..A... ........C. .T........ A..G......
DL54            .C.....T.. T..A..A... ........T. .C........ G..A......

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
Consensus       CCGGATGTTT TACAAACCAG CATCCGTAGC CTTATYGGCA AYCTYCTCTC
DL1 I           .......... .......... .......... .....T.... .C..T.....
DL2 I           .......... .......... .......... .....T.... .C..T.....
DL13 I          .......... .......... .......... .....T.... .C..T.....
DL16 I          .......... .......... .......... .....T.... .C..T.....
ST4 I           .......... .......... .......... .....T.... .C..C.....
R17 I           .......... .......... .......... .....C.... .T..T.....
J20 I           .......... .......... .......... .....T.... .C..T.....
MS2 I           .......... .......... .......... .....T.... .C..C.....
```

```
DL52          .......... .......... .......... .....T.... .C..T.....
DL54          .......... .......... .......... .....T.... .C..C.....
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
Consensus       TGGCTAYCRH TCGTCGTTGT TTGGGCAATG CACGTTYTCC AACGGTGCYY
DL1 I           ......T.GT .......... .......... ......C... ........CT
DL2 I           ......C.GC .......... .......... ......C... ........CT
DL13 I          ......C.GC .......... .......... ......C... ........CT
DL16 I          ......C.GC .......... .......... ......C... ........CT
ST4 I           ......C.GA .......... .......... ......C... ........CT
R17 I           ......C.AA .......... .......... ......C... ........CT
J20 I           ......T.GT .......... .......... ......T... ........CT
MS2 I           ......C.GA .......... .......... ......C... ........TC
DL52            ......C.GC .......... .......... ......C... ........CT
DL54            ......T.GT .......... .......... ......C... ........CT


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
Consensus       CDATGGGGCA CAAGTTGCAG GATGCAGCGC CTTACAAGAA GTTCGCTGAA
DL1 I           .G........ .......... .......... .......... ..........
DL2 I           .A........ .......... .......... .......... ..........
DL13 I          .A........ .......... .......... .......... ..........
DL16 I          .A........ .......... .......... .......... ..........
ST4 I           .T........ .......... .......... .......... ..........
R17 I           .T........ .......... .......... .......... ..........
J20 I           .T........ .......... .......... .......... ..........
MS2 I           .T........ .......... .......... .......... ..........
DL52            .A........ .......... .......... .......... ..........
DL54            .G........ .......... .......... .......... ..........


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
Consensus       CAAGCAACCG TTACCCCCCG CGCTCTRAGA GCGGCNYTAY TGGTCMGAGA
DL1 I           .......... .......... ......A... .....CC..C .....C....
DL2 I           .......... .......... ......G... .....AC..C .....C....
DL13 I          .......... .......... ......G... .....AC..C .....C....
DL16 I          .......... .......... ......G... .....AC..C .....C....
ST4 I           .......... .......... ......G... .....TC..T .....C....
R17 I           .......... .......... ......G... .....TC..T .....C....
J20 I           .......... .......... ......G... .....GT..C .....A....
MS2 I           .......... .......... ......G... .....TC..T .....C....
DL52            .......... .......... ......G... .....AC..C .....C....
DL54            .......... .......... ......G... .....TC..C .....C....


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
Consensus       CCARTGTGYG CCGTGGATYA GACACGCGGT CCRCTAYARY GARTCATATR
DL1 I           ...G....T. ........T. .......... ..G...C.AT ..G......G
DL2 I           ...G....T. ........C. .......... ..G...C.AC ..G......G
DL13 I          ...G....T. ........C. .......... ..G...C.AC ..G......G
DL16 I          ...G....T. ........C. .......... ..G...C.AC ..G......G
ST4 I           ...A....C. ........C. .......... ..G...T.AC ..G......G
R17 I           ...A....C. ........C. .......... ..A...T.AC ..G......G
J20 I           ...G....C. ........C. .......... ..G...T.AC ..A......A
MS2 I           ...A....C. ........C. .......... ..G...T.AC ..G......G
```

```
DL52          ...G....T. ........C. .......... ..G...C.AC ..G......G
DL54          ...G....C. ........C. .......... ..A...C.GC ..G......G

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                             560        570        580        590        600
Consensus     AATTTAGRCT CGTYGTAGGG AACGGAGTGT TYACAGTTCC GAAGAATAAT
DL1 I         .......G.. ...T...... .......... .T........ ..........
DL2 I         .......G.. ...C...... .......... .T........ ..........
DL13 I        .......G.. ...C...... .......... .T........ ..........
DL16 I        .......G.. ...C...... .......... .T........ ..........
ST4 I         .......G.. ...T...... .......... .T........ ..........
R17 I         .......G.. ...T...... .......... .T........ ..........
J20 I         .......A.. ...T...... .......... .T........ ..........
MS2 I         .......G.. ...T...... .......... .T........ ..........
DL52          .......G.. ...C...... .......... .T........ ..........
DL54          .......G.. ...C...... .......... .C........ ..........


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                             610        620        630        640        650
Consensus     AAAATAGATC GGGCTGCCTG YAAGGAGCCH GATRYGAAYA TGTACCTYCA
DL1 I         .......... .......... T........C ...AT...T. .......T..
DL2 I         .......... .......... T........C ...AT...T. .......C..
DL13 I        .......... .......... T........C ...AT...T. .......C..
DL16 I        .......... .......... T........C ...AT...T. .......C..
ST4 I         .......... .......... T........T ...AT...T. .......C..
R17 I         .......... .......... T........T ...AT...T. .......C..
J20 I         .......... .......... T........C ...AT...T. .......T..
MS2 I         .......... .......... T........T ...AT...T. .......C..
DL52          .......... .......... C........T ...GC...C. .......T..
DL54          .......... .......... T........A ...GC...C. .......T..


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                             660        670        680        690        700
Consensus     GAAAGGRGTY GGYGVYTTYA THMGVCGYCG BCTCMRHTCY RTYGGTATMG
DL1 I         ......G..T ..C.CC..C. .TA.A..C.. C...AAA..T G.C.....A.
DL2 I         ......A..T ..C.CC..C. .AA.G..T.. C...AAA..T G.C.....A.
DL13 I        ......A..T ..C.CC..C. .AA.G..T.. C...AAA..T G.C.....A.
DL16 I        ......A..T ..C.CC..C. .AA.G..T.. C...AAA..T G.C.....A.
ST4 I         ......G..C ..T.CC..T. .CA.A..C.. G...AAA..C G.T.....A.
R17 I         ......G..C ..C.CT..T. .TA.A..C.. G...AAA..C G.T.....A.
J20 I         ......A..T ..C.CT..T. .AC.G..T.. T...AGA..C G.T.....A.
MS2 I         ......G..C ..T.CT..C. .CA.A..C.. G...AAA..C G.T.....A.
DL52          ......A..T ..C.AT..T. .AC.C..T.. T...CGC..T A.C.....C.
DL54          ......G..T ..T.GC..C. .AC.C..C.. C...CGT..T A.C.....C.

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                             710        720        730        740        750
Consensus     AYCTGAATGA TCARWCGATC AAYCARCKTY TDGCWMARCW NGGCAGYRYM
DL1 I         .C........ ...AA..... ..T..G.G.C .A..TC.A.A G.....CATA
DL2 I         .C........ ...AA..... ..T..G.G.T .A..TC.A.A G.....CATA
DL13 I        .C........ ...AA..... ..T..G.G.T .A..TC.A.A G.....CATA
DL16 I        .C........ ...AA..... ..T..G.G.T .A..TC.A.A G.....CATA
ST4 I         .T........ ...AT..... ..C..G.T.C .G..TC.G.A G.....CGTA
R17 I         .C........ ...GT..... ..C..G.G.C .A..TC.G.A A.....CGCA
```

```
J20 I       .C........ ...AA..... ..C..G.G.C .G..TC.A.A A.....CGTC
MS2 I       .C........ ...AT..... ..C..G.G.C .G..TC.G.A G.....CGTA
DL52        .C........ ...AA..... ..T..G.G.C .T..AA.A.T T.....CATC
DL54        .C........ ...AA..... ..T..A.G.C .T..AA.A.T C.....TATC
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        760        770        780        790        800
Consensus       GATGGYTCDY TDGCRACGAT WGAYYTATCG TCNGCNTCBG AYTCYATHTC
DL1 I           .....T..TT .A..G..... A..CC..... ..A..T..G. .T..C..C..
DL2 I           .....C..TT .A..A..... A..TT..... ..G..T..T. .T..T..C..
DL13 I          .....C..TT .A..A..... A..TT..... ..G..T..T. .T..T..C..
DL16 I          .....C..TT .A..A..... A..TT..... ..G..T..T. .T..T..C..
ST4 I           .....T..GC .T..G..... A..CT..... ..T..A..C. .T..C..C..
R17 I           .....T..GC .T..G..... A..CT..... ..C..G..C. .T..C..T..
J20 I           .....C..AT .G..G..... A..CT..... ..G..C..T. .C..C..C..
MS2 I           .....T..GC .T..G..... A..CT..... ..T..A..C. .T..C..C..
DL52            .....T..TT .A..G..... T..TT..... ..T..T..T. .T..C..A..
DL54            .....T..TT .A..G..... A..TT..... ..T..T..T. .C..C..A..


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        810        820        830        840        850
Consensus       BGAYCGCCTN GTKTGGRRNT TYCTYCCAYC KSARHTRTAY KCRTAYCTBK
DL1 I           T..C.....G ..G...AGC. .T..C...C. TG.GC.A..T T.A..T..CG
DL2 I           T..C.....G ..G...AGT. .C..C...C. TG.GT.A..T T.A..T..CG
DL13 I          T..C.....G ..G...AGT. .C..C...C. TG.GT.A..T T.A..T..CG
DL16 I          T..C.....G ..G...AGT. .C..C...C. TG.GT.A..T T.A..T..CG
ST4 I           C..T.....G ..G...AGT. .T..C...C. TG.GC.A..T T.A..T..CG
R17 I           C..C.....A ..G...AAT. .C..T...C. TG.GC.A..T T.A..T..TG
J20 I           T..C.....G ..G...AGT. .C..C...C. TG.GC.A..T T.A..T..CG
MS2 I           C..T.....G ..G...AGT. .T..C...C. TG.GC.A..T T.A..T..CG
DL52            G..C.....C ..T...GAA. .C..C...T. GC.AA.G..C G.G..C..GT
DL54            G..C.....T ..T...GAG. .T..C...C. GC.AA.G..T G.G..C..GT


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        860        870        880        890        900
Consensus       MKMRWATYCG YTCVYMMYRH GGAATCRTWG AYGGVSRKRY RRTMSRVTGG
DL1 I           ATCGT..C.. C..CCACTAC ......G.A. .T..CGAGAC GA.ACGG...
DL2 I           ATCGT..C.. C..CCACTAC ......G.A. .T..GGAGAC GA.ACGA...
DL13 I          ATCGT..C.. C..CCACTAC ......G.A. .T..GGAGAC GA.ACGA...
DL16 I          ATCGT..C.. C..CCACTAC ......G.A. .T..GGAGAC GA.ACGA...
ST4 I           ATCGT..C.. C..ACACTAC ......G.A. .T..CGAGAC GA.ACGA...
R17 I           ATCGT..C.. C..GCACTAC ......G.A. .T..CGAGAC GA.ACGA...
J20 I           ATCGT..C.. C..CCACTAT ......A.A. .T..AGAGAC GA.ACGA...
MS2 I           ATCGT..C.. C..ACACTAC ......G.A. .T..CGAGAC GA.ACGA...
DL52            CGAAA..T.. T..GTCACGA ......G.T. .C..ACGTGT AG.CGAC...
DL54            CTAAA..T.. T..GCCACGA ......G.T. .C..ACGTAT GA.CGAC...


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        910        920        930        940        950
Consensus       SAMCTATTTT CCACDATGGG WAATGGDTTY ACNTTYGARC TAGAGTCCAT
DL1 I           G.A..... ....A..... A.....G..C ..T..C..A. ..........
DL2 I           G.A..... ....A..... T.....G..C ..T..C..A. ..........
DL13 I          G.A..... ....A..... T.....G..C ..T..C..A. ..........
DL16 I          G.A..... ....A..... T.....G..C ..T..C..A. ..........
ST4 I           G.A..... ....A..... A.....G..C ..G..T..G. ..........
R17 I           G.A..... ....G..... A.....G..T ..A..T..G. ..........
J20 I           G.A..... ....A..... A.....G..C ..T..T..A. ..........
MS2 I           G.A..... ....A..... A.....G..C ..A..T..G. ..........
```

DL52 C.C..... ....A..... T.....T..T ..C..T..G. ....
DL54 C.C..... ....T..... T.....A..T ..C..T..A. ....

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       960        970        980        990       1000
Consensus       GATMTTYTGG GCWATAGTNA ARGCRACYMW RAYYCATTTT GGTAACSYYG
DL1 I           ...A..T... ..T.....G. .A..G..CCA G.TC...... ......GCC.
DL2 I           ...C..T... ..T.....G. .A..A..CCA G.TC...... ......GCC.
DL13 I          ...C..T... ..T.....G. .A..A..CCA G.TC...... ......GCC.
DL16 I          ...C..T... ..T.....G. .A..A..CCA G.TC...... ......GCC.
ST4 I           ...A..C... ..A.....C. .A..G..CCA A.TC...... ......GCC.
R17 I           ...A..C... ..A.....C. .A..A..CCA A.TC...... ......GCC.
J20 I           ...A..T... ..T.....A. .A..A..CCA G.TC...... ......GCC.
MS2 I           ...A..C... ..A.....C. .A..G..CCA A.TC...... ......GCC.
DL52            ...A..T... ..A.....C. .G..A..TAT G.TC...... ......CTT.
DL54            ...C..T... ..A.....T. .G..A..CAT G.CT...... ......CTT.

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1010       1020       1030       1040       1050
Consensus       GAACMATWGG CATCTAYGGG GACGATATYA TATGTCCCAS WGAGATTGCA
DL1 I           ....C..A.. ......C... ........T. .........G T.........
DL2 I           ....C..A.. ......C... ........T. .........G T.........
DL13 I          ....C..A.. ......C... ........T. .........G T.........
DL16 I          ....C..A.. ......C... ........T. .........G T.........
ST4 I           ....C..A.. ......C... ........T. .........G T.........
R17 I           ....C..A.. ......C... ........C. .........G T.........
J20 I           ....C..A.. ......C... ........T. .........G T.........
MS2 I           ....C..A.. ......C... ........T. .........G T.........
DL52            ....A..T.. ......T... ........C. .........C A.........
DL54            ....A..T.. ......T... ........C. .........C A.........

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1060       1070       1080       1090       1100
Consensus       CCYCGTGTGC TRGAGGCDCT HRSCTWCTAC GGTTTYAAAC CGAATCWBYS
DL1 I           ..C....... .A.....A.. TGC..A.... .....T.... ......TTCG
DL2 I           ..C....... .A.....A.. TGC..A.... .....C.... ......TTCG
DL13 I          ..C....... .A.....A.. TGC..A.... .....C.... ......TTCG
DL16 I          ..C....... .A.....A.. TGC..A.... .....C.... ......TTCG
ST4 I           ..C....... .G.....A.. TGC..A.... .....C.... ......TCCG
R17 I           ..C....... .A.....A.. TGC..A.... .....T.... ......TTCG
J20 I           ..C....... .A.....G.. AGC..A.... .....C.... ......TTCG
MS2 I           ..C....... .A.....A.. TGC..A.... .....T.... ......TTCG
DL52            ..T....... .A.....T.. CAG..T.... .....T.... ......AGTC
DL54            ..T....... .A.....T.. CAG..T.... .....T.... ......AGTC

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1110       1120       1130       1140       1150
Consensus       KAARACGTTC RTSWCVGGKC KCTTTCGCGA GWSCTGYRGY GCRCAYTWYT
DL1 I           T..A...... G.GT.G..G. T......... .AG...CG.C ..G..C.TT.
DL2 I           T..A...... G.GT.A..G. T......... .AG...TG.C ..G..C.TT.
DL13 I          T..A...... G.GT.A..G. T......... .AG...TG.C ..G..C.TT.
DL16 I          T..A...... G.GT.A..G. T......... .AG...TG.C ..G..C.TT.
ST4 I           T..A...... G.GT.C..G. T......... .AG...CG.C ..G..C.TT.
R17 I           T..A...... G.GT.C..G. T......... .AG...CA.C ..G..C.TT.
J20 I           T..A...... G.GT.A..G. T......... .AG...TG.C ..G..C.TT.
MS2 I           T..A...... G.GT.C..G. T......... .AG...CG.C ..G..C.TT.
```

```
DL52          G..G..... A.CA.G..T. G......... .TC...TG.T ..A..T.AC.
DL54          G..G..... A.CA.G..T. G......... .TC...TG.T ..A..T.AC.

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                         1160       1170       1180       1190       1200
Consensus     WCSGYGGTGY YGATKKCAAA CCGWTYTAYA TCARGAAACC HGTYRACAAY
DL1 I         A.C.T....T C...GT.... ...T.T..C. ...A...... A..TG....C
DL2 I         A.C.T....T C...GT.... ...T.C..C. ...A...... A..TG....T
DL13 I        A.C.T....T C...GT.... ...T.C..C. ...A...... A..TG....T
DL16 I        A.C.T....T C...GT.... ...T.C..C. ...A...... A..TG....T
ST4 I         A.C.T....T C...GT.... ...T.T..C. ...A...... T..TG....T
R17 I         A.C.T....T C...GT.... ...T.T..C. ...A...... T..TG....C
J20 I         A.C.T....T C...GT.... ...T.T..C. ...G...... A..CG....T
MS2 I         A.C.T....T C...GT.... ...T.T..C. ...A...... T..TG....T
DL52          T.G.C....C T...TG.... ...A.T..T. ...A...... C..TA....C
DL54          T.G.C....C T...TG.... ...A.T..T. ...A...... C..TA....C


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                         1210       1220       1230       1240       1250
Consensus     CTCTTYKCCS TBWKKCTGWT MHTNAAYMGR CTDMGSGGNT GGGGNGTTGT
DL1 I         .....CT..C .TATG...A. AA.G..TC.G ..TA.G..G. ....C.....
DL2 I         .....CT..C .TATG...A. CA.G..TC.G ..TA.G..G. ....C.....
DL13 I        .....CT..C .TATG...A. CA.G..TC.G ..TA.G..G. ....C.....
DL16 I        .....CT..C .TATG...A. CA.G..TC.G ..TA.G..G. ....C.....
ST4 I         .....CG..C .TATG...A. AT.G..TC.G ..AC.G..T. ....A.....
R17 I         .....TG..C .GATG...A. AT.A..TC.G ..AC.G..T. ....G.....
J20 I         .....CT..C .TATG...A. CA.G..TC.G ..TA.G..A. ....T.....
MS2 I         .....CG..C .GATG...A. AT.A..TC.G ..AC.G..T. ....A.....
DL52          .....CG..G .CTGT...T. AC.C..CA.G ..AC.C..G. ....T.....
DL54          .....CG..G .CTGT...T. AC.T..CA.A ..GC.C..C. ....T.....


                    ....|....| ....|....| ....|....| ....|....| ....|....|
                         1260       1270       1280       1290       1300
Consensus     SRRMGGWRTG TCAGATCCDC GCCTHTWYRA GRYKTGGRWA YGRCTVTCCK
DL1 I         CGGA..TA.. ........G. ....T.ATA. .GTT...GT. C.A..G...T
DL2 I         CGGA..TA.. ........G. ....T.ATA. .GTT...GT. C.A..A...T
DL13 I        CGGA..TA.. ........G. ....T.ATA. .GTT...GT. C.A..A...T
DL16 I        CGGA..TA.. ........G. ....T.ATA. .GTT...GT. C.A..A...T
ST4 I         CGGA..TA.. ........A. ....T.ACA. .GTG...GT. C.A..C...T
R17 I         CGGA..TA.. ........A. ....T.ACA. .GTG...GT. C.A..C...T
J20 I         CGGA..TA.. ........T. ....A.ATA. .GTT...GT. C.A..G...T
MS2 I         CGGA..TA.. ........A. ....C.ATA. .GTG...GT. C.G..C...T
DL52          GAAC..AG.. ........T. ....C.TCG. .ACT...AA. T.G..A...G
DL54          GAAC..TG.. ........T. ....C.TCG. .ACT...AA. T.G..A...G

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                         1310       1320       1330       1340       1350
Consensus     MMCDKGTDCC HTCGATRYTY TTCGGTGGGW CDRACCTYGM TGCHGACTAY
DL1 I         CC.TG..T.. A.....GT.C .........A .GG....T.C ...C.....T
DL2 I         CC.TG..T.. A.....GT.C .........A .GG....T.C ...C.....C
DL13 I        CC.TG..T.. A.....GT.C .........A .GG....T.C ...C.....C
DL16 I        CC.TG..T.. A.....GT.C .........A .GG....T.C ...C.....C
ST4 I         CC.AG..G.. T.....GT.T .........A .GG....C.C ...C.....C
R17 I         CC.AG..A.. T.....GT.C .........A .GG....C.C ...C.....C
```

```
J20 I       CC.TG..T.. A.....GT.C .........A .GG....T.C ...T.....C
MS2 I       CC.AG..G.. T.....GT.C .........A .GG....C.C ...C.....C
DL52        AA.GT..T.. T.....AC.T .........T .TA....T.A ...A.....C
DL54        AA.GT..T.. C.....AC.T .........T .AA....T.A ...A.....T
```

First nt insertion

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1360       1370       1380       1390       1400
Consensus        TACGTRGTYA GYCSBCMBRM NGCYADRWMK CRKTDYGABR VYMANMCHBY
DL1 I            .....A..C. .C.CC.CGAC C..--AGTCT .AG.TT-.TA CTA.AA.TGC
DL2 I            .....A..T. .C.CT.CGAA T..--TGTCT .GG.TT-.TA CCA.AA.TGC
DL13 I           .....A..T. .C.CT.CGAA T..--TGTCT .GG.TT-.TA CCA.AA.TGC
DL16 I           .....A..T. .C.CT.CGAA T..--TGTCT .GG.TT-.TA CCA.AA.TGC
ST4 I            .....A..C. .C.CG.CCAC G..--AGTCT .GG.AT-.TA CCA.GA.TCC
R17 I            .....A..C. .C.CG.CTAC G..--GGTCT .GG.AT-.TA CCA.GA.TCC
J20 I            .....A..C. .C.CC.CGAC A..--GGTCT .AG.TT-.TA CTA.GA.CGC
MS2 I            .....A..C. .C.CG.CTAC G..--AGTCT .GG.AT-.CA CCA.GA.TCC
DL52             .....G..T. .T.GT.AGGA A..T.AAAAG .GT.AC..GA GTC.CC.ATC
DL54             .....G..C. .C.GC.AGGA A..C.GGAAG .GT.GC..GG ATC.TC.AGT

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1410       1420       1430       1440       1450
Consensus        NTAYGGRMGB HYVCTYKCKS AYDYYCRYAS CWCKSSTYWY CGKYTKKMYC
DL1 I            C..C..GA.G TTA..CG.GG .TACC.GT.C .T.GGG.TTC ..TC.TGCT.
DL2 I            A..T..GA.G TTA..CG.GG .CGCC.GT.C .T.GGG.TTC ..TC.TGCT.
DL13 I           A..T..GA.G TTA..CG.GG .CGCC.GT.C .T.GGG.TTC ..TC.TGCT.
DL16 I           A..T..GA.G TTA..CG.GG .CGCC.GT.C .T.GGG.TTC ..TC.TGCT.
ST4 I            G..T..GC.G CTA..CG.GG .TACC.GT.C .T.GGG.TTC ..TC.TGCT.
R17 I            G..T..AC.G CTG..CG.GG .TACC.GT.C .T.GGG.TTC ..TC.TGCT.
J20 I            A..T..GA.G TTG..CG.GG .CGCC.GT.C .T.GGG.TTC ..TC.TGCT.
MS2 I            G..C..GC.G CTG..CG.GG .TACC.GT.C .T.GGG.TTC ..TC.TGCT.
DL52             T..T..GC.C ACC..CT.TC .CTTT.AT.G .A.TCC.CAT ..GC.TTAC.
DL54             A..T..AC.T ACC..TG.TC .TTTC.AC.G .T.TCC.CAT ..GT.GTAC.

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1460       1470       1480       1490       1500
Consensus        GYRYRCCCGY NHVWRARCGH RARYDYTCTY MRCGHRAAGM RYSRCAGYGG
DL1 I            .TAT----.C AAAAG.G..A A.GCGC.-.T AG..AG...C ATGA...T..
DL2 I            .TAT----.C GAAAG.G..A A.GCAC.-.T AG..AG...C ATGA...T..
DL13 I           .TAT----.C GAAAG.G..A A.GCAC.-.T AG..AG...C ATGG...T..
DL16 I           .TAC----.C GAAAG.G..A A.GCAC.-.T AG..AG...C ATGA...T..
ST4 I            .TAT----.C TCGAG.A..C A.GTTC.-.C AG..AA...C ATGA...T..
R17 I            .TAT----.C TCGAG.A..C A.GTTC.-.C AG..AA...C ATGA...T..
J20 I            .TAT----.C GAAAG.G..T A.GCAC.-.T AG..AG...C ATGA...T..
MS2 I            .TAT----.C TCGAG.A..C A.GTTC.-.C AG..AA...C ACGA...T..
DL52             .CGTG....T ATCTA.A..T G.ATTT...C CA..TG...A ATC----C..
DL54             .CGTA....C CTCTA.A..C G.ATTT...T CG..CG...A GTC----C..

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         1510       1520       1530       1540       1550
Consensus        YCGCYWCATM RCDTGGTWCC ATAMTGGHGG TSARRTYAYY GAYASYAYKA
DL1 I            T...TA...A G.A....T.. ...C...A.. .G.GA.C.CC ..T.GT.TG.
DL2 I            T...TA...A G.A....T.. ...C...A.. .G.AA.C.CC ..T.GT.TG.
DL13 I           T...TA...A G.A....T.. ...C...A.. .G.AA.C.CC ..T.GT.TG.
DL16 I           T...TA...A G.A....T.. ...C...A.. .G.AA.C.CC ..T.GT.TG.
ST4 I            C...TA...A G.G....T.. ...C...A.. .G.AG.C.CC ..C.GT.TG.
R17 I            T...TA...A G.G....T.. ...C...A.. .G.AA.C.CC ..C.GT.TG.
J20 I            T...TA...A G.A....T.. ...C...A.. .G.AA.C.CC ..T.GT.TG.
MS2 I            T...TA...A G.G....T.. ...C...A.. .G.AA.C.CC ..C.GC.TG.
DL52             T...CT...C A.T....A.. ...A...T.. .C.AG.T.TT ..C.CC.CT.
```

DL54 T...CT...C A.T....A.. ...A...C.. .C.AA.T.TT ..C.CC.CT.

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1560       1570       1580       1590       1600
Consensus       MGWCCSCHRG SGTKCGYVTH RTRCGMACDT CGGARTGGCT RRCRSYRGTK
DL1 I           A.T..G.TG. C..G..CG.A A.G..C..T. ....G..... GA.GCCG..T
DL2 I           A.T..G.TG. C..G..TG.A A.A..C..T. ....G..... GA.GCCG..T
DL13 I          A.T..G.TG. C..G..TG.A A.A..C..T. ....G..... GA.GCCG..T
DL16 I          A.T..G.TG. C..G..TG.A A.A..C..T. ....G..... GA.GCCG..T
ST4 I           A.T..G.CG. C..G..TA.T A.G..C..T. ....G..... AA.GCCG..T
R17 I           A.T..G.CG. C..G..CA.C A.G..C..T. ....G..... AA.GCCG..T
J20 I           A.T..G.CG. C..G..TG.T A.G..C..T. ....A..... AA.GCCG..T
MS2 I           A.T..G.CG. C..G..CG.T A.A..C..T. ....G..... AA.GCCG..T
DL52            C.A..C.AA. G..T..CC.T G.G..A..G. ....A..... AA.AGTG..G
DL54            C.A..C.AA. G..T..CC.T G.G..A..A. ....A..... AG.GGTA..G

                                                      (Group I) ORF4 STOP
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1610       1620       1630       1640       1650
Consensus       CCMHYMTTCC CKCAGGARKR TGRCRMMHGC GAGCTCTCCT MGKYAGCWSR
DL1 I           ..CACA.... .T.....GTG ..G-GCCA.. .......... C.GT...TGA
DL2 I           ..CATA.... .T.....ATG ..G-GCCA.. .......... C.GT...TGA
DL13 I          ..CATA.... .T.....ATG ..G-GCCA.. .......... C.GT...TGA
DL16 I          ..CATA.... .T.....ATG ..G-GCCA.. .......... C.GT...TGA
ST4 I           ..CACA.... .T.....GTG ..G-GCCA.. .......... C.GT...TGA
R17 I           ..CACA.... .T.....GTG ..G-GCCA.. .......... C.GT...TGA
J20 I           ..CACA.... .G.....GTG ..G-GCCA.. .......... C.GT...TGA
MS2 I           ..CACA.... .T.....GTG ..G-GCCA.. .......... C.GT...TGA
DL52            ..ACTC.... .T.....AGA ..G.AACT.. .......... A.TC...ACG
DL54            ..ATCC.... .T.....AGA ..A.ACAC.. .......... A.TC...ACG

              ORF4 STOP (JS Group)
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1660       1670       1680       1690       1700
Consensus       SCKAGGGACC CCCGTAAWCG GGGTGGGTGT GCTCGAAAGA GCACGGGTCC
DL1 I           C.G....... .......A.. .......... .......... ..........
DL2 I           C.G....... .......A.. .......... .......... ..........
DL13 I          C.G....... .......A.. .......... .......... ..........
DL16 I          C.G....... .......A.. .......... .......... ..........
ST4 I           C.G....... .......A.. .......... .......... ..........
R17 I           C.G....... .......A.. .......... .......... ..........
J20 I           C.G....... .......A.. .......... .......... ..........
MS2 I           C.G....... .......A.. .......... .......... .........--
DL52            G.T....... .......T.. .........- ---------- --------..
DL54            G.T....... .......T.. .........- ---------- --------..

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       1710       1720       1730       1740       1750
Consensus       GCGWAAGCGG YCCGKMTYCA YCGWAAGGTG GGCGGGCTTC GGCCCARSGA
DL1 I           ...A...... T--.GC.C.. C..A...... .......... ......GG..
DL2 I           ...A...... T--.GC.C.. C..A...... .......... ......GG..
DL13 I          ...A...... T--.GC.C.. C..A...... .......... ......GG..
DL16 I          ...A...... T--.GC.C.. C..A...... .......... ......GG..
ST4 I           ...A...... T--.GC.C.. C..A...... .......... ......GG..
R17 I           ...A...... T--.GC.C.. C..A...... .......... ......GG..
J20 I           ...T...... T--.GC.C.. C..A...... .......... ......GG..
MS2 I           ...A...... T...GC.C.. C..A...... .......... ......GG..
DL52            ...A...... C---TA.T.. T..T..---- ---------- -----.AC..
```

DL54 ...A..... C---TA.T.. T..T..---- ---------- -----..AC..

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        1760       1770       1780       1790       1800
Consensus       YCTCCCCYTR ARGAKAGGRC CCGGKAWYCT CCYRATTYGG TRACTARCTT
DL1 I           C......C.A .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
DL2 I           C......C.A .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
DL13 I          C......C.A .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
DL16 I          C......C.A .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
ST4 I           C......T.G .A..G...G. ....G.TT.. ..CG...T.. .A....G..-
R17 I           C......C.G .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
J20 I           C......C.A .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
MS2 I           C......C.A .A..G...A. ....G.TT.. ..CG...T.. .A....G..-
DL52            T--------- -G..T...A- ---.T.AT.. ..TA...T.. .G....A...
DL54            T--------- -G..T...A- ---.T.AC.. ..TA...C.. .G....A...


                ....|....| ....|....| .
                        1810       1820
Consensus       TGCTTGGCTA GTBACCACCC A
DL1 I           -......... ..T....... .
DL2 I           -......... ..T....... .
DL13 I          -......... ..T....... .
DL16 I          -......... ..T....... .
ST4 I           -......... ..T....... .
R17 I           -......... ..G....... .
J20 I           -......... ..T....... .
MS2 I           -......... ..T....... .
DL52            .......... ..C....... .
DL54            .......... ..C....... .
```